

Bibliography

[1] Van der Vaart, A. W. (2000). Asymptotic statistics (Vol. 3). Cambridge university press.

[2] 김우철, 수리통계학.

[3] 김우철, 통계이론¹ 강의노트.

Little o

(정의) $\{x_n\}, \{r_n\}$ 은 실수열이라고 하자.

$$\begin{aligned} x_n &= o(r_n) \quad \text{as } n \rightarrow \infty \\ \iff x_n/r_n &\rightarrow 0 \quad \text{as } n \rightarrow \infty \end{aligned}$$

note: $x_n = o(1)$ 의 의미는 $x_n \rightarrow 0$ 의 의미이다.

note: $x_n = o(r_n)$ 의 의미는 $x_n/r_n \rightarrow 0$ 의 의미이다.

example: $x_n = 1/n$ 이면 $x_n = o(1)$.

example: $x_n = 1/n^2$ 이면 $x_n = o(1/n)$

note: $x_n = 1/n^2$ 일 경우 $x_n = o(1), x_n = o(1/\sqrt{n}), x_n = o(1/n\sqrt{n})$ 도 가능함.

example: $x_n = n^2$ 이면 $x_n = o(n^3)$.

(느낌) $x_n = o(\star_n)$ 의 표현은 x_n 이 \star_n 보다 빠르게 0으로 수렴 (혹은 천천히 ∞ 로 발산) 하기만 하면 된다.



(정의) $f(x), g(x)$ 는 실변수 함수

$$\begin{aligned} f(x) &= o(g(x)) \quad \text{as } x \rightarrow a \\ \iff f(x)/g(x) &\rightarrow 0 \quad \text{as } x \rightarrow a \end{aligned}$$

(느낌) 함수열의 경우도 동일하게 해석할 수 있다.

$$f(x) = o(\star(x)) \quad x \rightarrow a$$

는 x 가 점점 a 로 가까워질때 $f(x)$ 가 $\star(x)$ 보다 빠르게 0으로 수렴하거나 천천히 ∞ 로 발산하면 된다.



Big O

(정의) $\{x_n\}, \{r_n\}$ 은 실수열이라고 하자.

$$\begin{aligned} x_n &= O(r_n) \quad \text{as } n \rightarrow \infty \\ \iff \sup_n |x_n/r_n| &< \infty \end{aligned}$$

example: $x_n = n^2 + 2n + 100$ 이면 $x_n = O(n^2)$.

example: $x_n = n^{-2} + n^{-1} + 1$ 이면 $x_n = O(1)$.

(느낌) 따라서

$$x_n = O(\star_n)$$

$\{x_n\}$ 의 최고차항이 \star_n 의 최고차항과 같다는 의미이다.



(정의) $f(x), g(x)$ 는 실변수함수 라고 하자.

$$\begin{aligned} f(x) &= O(g(x)) \quad as \ x \rightarrow a \\ \iff f(x)/g(x) &\rightarrow 0 \quad as \ x \rightarrow a \end{aligned}$$



Little o_p

(정의) $\{X_n\}, \{R_n\}$ 은 확률변수열이라고 하자.

$$\begin{aligned} X_n &= o_p(R_n) \quad as \ n \rightarrow \infty \\ \iff X_n/R_n &\xrightarrow{p} 0 \quad as \ n \rightarrow \infty \end{aligned}$$

note: $X_n = o_p(R_n)$ 의 의미는 새로운 확률변수 $Y_n = X_n/R_n$ 의 수열이 $\{Y_n\}$ 이 0으로 확률수렴한다는 의미이다.

note: 여기에서 $\{R_n\}$ 은 확률변수열이 아니라 실수열로 생각할 수 있다. (실수는 확률변수의 한 종류이므로)

note: $\{X_n\}$ 과 $\{R_n\}$ 이 모두 실수열이라면 O_p 의 정의가 O 의 정의와 같아진다.



Big O_p

(정의) $\{X_n\}, \{R_n\}$ 은 확률변수열이라고 하자.

$$\begin{aligned} X_n &= O_p(R_n) \quad as \ n \rightarrow \infty \\ \iff \sup_n P\{|X_n/R_n| > M\} &\xrightarrow{M \rightarrow \infty} 0 \end{aligned}$$

- 이게 원래 실수버전으로 생각하면 엄청 큰 M 에 대해서

$$\sup_n |X_n/R_n| < M < \infty$$

이어야 한다. 그런데 X_n/R_n 은 확률변수이므로 위의 부등식이 성립하지 않게 된다. 결국 M 을 엄청 크게 잡으면

$$\sup_n |X_n/R_n| > M$$

이 될 확률은 너무 작지않겠냐? 라는 식의 논의를 해야한다. 즉 아래의 느낌이 되어야 한다.

$$\text{for very large } M : \quad P\{\sup_n |X_n/R_n| > M\} \approx 0$$

- 이런 느낌을 살린것이 바로 아래의 statement이다.

$$P\{\sup_n |X_n/R_n| > M\} \xrightarrow{M \rightarrow \infty} 0$$

(느낌) 결국 $X_n = O_p(1)$ 의 의미는

$$\text{for very large } M : \quad P\{\sup_n |X_n| < M\} \approx 1$$

의 의미가 된다. 즉 확률변수 X_n 이 어딘가에 bound 되어 있다는 의미가 된다.

(정리) $Y_n \xrightarrow{d} Y$ 이라면 $Y_n = O_p(1)$ 이다.

note: 이 정리는 김우철 수리통계학 [2] 연습문제 5.11, 김우철 통계이론 강의노트 [3] 의 Review C, 그 외 다수에 소개되어 있다.

(정의) (*tightness of X*)

The random variable X is *tight*

$$\stackrel{\text{def}}{\iff} \forall \epsilon > 0 \quad \exists M \quad \text{s.t.} \quad P(|X| > M) < \epsilon$$

$$\stackrel{\text{def}}{\iff} P(|X| > M) \rightarrow 0 \quad \text{as } M \rightarrow \infty$$

$$\stackrel{\text{def}}{\iff} \mu^*([-M, M]^c) \rightarrow 0 \quad \text{as } M \rightarrow \infty$$

$$\stackrel{\text{def}}{\iff} \epsilon > 0 \quad \exists M \quad \text{s.t.} \quad 1 - F(M) + F(-M) < \epsilon$$

note: $\mu^*([0, M)) = F(M) = P(X < M)$

note: $\mu^*([-M, M]^c) = 1 - (P(X < M) - P(X < -M)) = 1 - F(M) + F(-M)$

note: 아무리 ϵ 을 작게 잡아도 결국 M 을 크게 잡으면 된다.

note: 따라서 대충 $M \approx \infty$ 일때 $P(|X| > M) \approx 0$ 인 느낌이다.

(정리) 모든 확률변수 X 는 *tight* 하다.

(정의) (*uniformly tightness of $\{X_n\}$*)

$\{X_n\}$ is *uniformly tight*

$$\stackrel{\text{def}}{\iff} \forall \epsilon > 0 \quad \exists M \quad \text{s.t.} \quad \sup_n P(|X_n| > M) < \epsilon$$

$$\stackrel{\text{def}}{\iff} \sup_n P(|X_n| > M) \rightarrow 0 \quad \text{as } M \rightarrow \infty$$

$$\stackrel{\text{def}}{\iff} \sup_n \mu_n^*([-M, M]^c) \rightarrow 0 \quad \text{as } M \rightarrow \infty$$

$$\stackrel{\text{def}}{\iff} \epsilon > 0 \quad \exists M \quad \text{s.t.} \quad \limsup_{n \rightarrow \infty} (1 - F_n(M) + F_n(-M)) < \epsilon$$

$$\stackrel{\text{def}}{\iff} X_n = O_p(1) \quad \text{as } n \rightarrow \infty$$

• 결국 확률변수열이 $\{X_n\}$ 이 *uniformly tight* 하다는 것은 확률적으로 유계라는 말과 같다. ([1] p.8.)

note: 애초에 *uniformly tightness* 는 $\{X_n\}$ 이 *iid* 인 상황에서는 살짝 어색한 개념이다. 왜냐하면 *iid* 인 경우라면 하나의 분포 X_1 만 잡아서 확률적으로 유계임을 보이면 되기 때문이다. 즉 *uniformly* 라는 말을 쓸 필요가 없다. 직관적으로 *uniformly*라는 말을 쓴다는 것은 sup을 취한다는 느낌을 받아야 하는데 *iid* 의 경우

$$\sup_n P(|X_n| > M) = P(|X_1| > M)$$

이 되어서 sup이 그냥 날아가버린다.

Calculus with O_p and o_p

(정리) 아래는 증명없이 받아들이자 [1]. 여기에서 $\{R_n\}$ 은 (아무런 추가 조건없는) 확률변수열이다.

(1) $O_p(1) + O_p(1) = O_p(1)$.

(2) $o_p(1) + o_p(1) = o_p(1)$.

(3) $o_p(1) + O_p(1) = O_p(1)$.

(4) $O_p(1)O_p(1) = O_p(1)$.

(5) $o_p(1)o_p(1) = o_p(1).$

(6) $O_p(1)o_p(1) = o_p(1).$

(7) $(1 + o_p(1))^{-1} = O_p(1).$

(8) $o_p(R_n) = R_n o_p(1).$

(9) $O_p(R_n) = R_n O_p(1).$

(10) $o_p(O_p(1)) = o_p(1)$

note: (2)번 공식, 즉 $o_p(1) + o_p(1) = o_p(1)$ 가 의미하는것은 $X_n \xrightarrow{p} 0$ 이고 $Y_n \xrightarrow{p} 0$ 일때 $Z_n = X_n + Y_n \xrightarrow{p} 0$ 라는 의미이다. 이는 continuous mapping thm 즉 김우철 정리 5.2.3 [2]의 한 형태가 된다.

note: (6)번 공식, 즉 $O_p(1)o_p(1) = o_p(1)$ 이 의미하는 것은 (1) $\{X_n\}$ 이 확률적으로 유계이고 (2) $\{Y_n\}$ 이 o으로 확률수렴한다면 (★) $\{X_n Y_n\}$ 은 0 으로 확률수렴한다는 의미이다. 여기에서 (1) 대신에 $\{X_n\}$ 이 어떠한 분포로 분포수렴한다는 가정이 추가적으로 있다면 이는 슬러츠키 정리의 한 형태가 된다.

Delta method

(정리) $\{X_n\}$ 이 확률변수열이라고 하자. (1) $\sqrt{n}(X_n - \theta) \xrightarrow{d} Z$ 이고 (2) $g(x)$ 가 θ 에서 미분가능하다면 아래가 성립한다.

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{d} \dot{g}(\theta)Z$$

note: $\{X_n\}$ 이 iid 일 필요는 없다.

note: 일반적인 함수 $g : \mathbb{R}^k \rightarrow \mathbb{R}^m$ 에 대하여 (2)의 조건은 아래와 같이 표현가능하다. [1]

(1) g is differentialbe at θ

(2) there is linear map (matrix) $g'_\theta : \mathbb{R}^k \rightarrow \mathbb{R}^m$ such that

$$g(\theta + h) - g(\theta) = g'_\theta(h) + o(\|h\|)$$

여기에서 linear map $h \rightarrow g'_\theta(h)$ 을 *total derivate* 라고 부른다. (이는 *partial derivate*와 구분된다.)

note: $g(x) : \mathbb{R}^k \rightarrow \mathbb{R}^m$ 이 *totally differentiable* 하기 해서는 (1) $g(x)$ 의 모든 편미분이 존재하고 즉 모든 $i = 1, 2, \dots, k$ and $j = 1, 2, \dots, m$ 에 대하여

$$\frac{\partial g_j(x)}{\partial x_i}$$

가 존재하고 (2) 그것이 연속이어야 한다. 단순히 편미분이 존재하기만 해서는 안된다.

note: 위의 조건을 *continuously differentiable* 이라고 하기도 한다.

• 증명을 해보자. 아래가 성립한다. $g(x)$ 를 $x = a$ 에서 테일러 전개하면

$$g(x) = g(a) + (x - a)\dot{g}(a) + O((x - a)^2) \quad as \ x \rightarrow a$$

그런데 $O((x - a)^2) = (x - a)^2 O(1)$ 이므로

$$g(x) = g(a) + (x - a)(\dot{g}(a) + (x - a)O(1)) \quad as \ x \rightarrow a$$

양변에 \sqrt{n} 을 곱하면

$$\sqrt{n}(g(x) - g(a)) = \sqrt{n}(x - a)(\dot{g}(a) + (x - a)O(1)) \quad as \ x \rightarrow a$$

- x 대신에 X_n 을 대입하고 a 대신에 θ 를 대입하면

$$\sqrt{n}(g(X_n) - g(\theta)) = \sqrt{n}(X_n - \theta)(\dot{g}(\theta) + (X_n - \theta)O(1)) \quad as \ X_n \rightarrow \theta$$

정리하면

$$\sqrt{n}(g(X_n) - g(\theta)) = \sqrt{n}(X_n - \theta)\dot{g}(\theta) + \sqrt{n}(X_n - \theta)^2O(1) \quad as \ X_n \rightarrow \theta$$

- 그런데 조건 (1) 에 따라서 $\sqrt{n}(X_n - \theta) = O_p(1)$ 이다. 따라서 $X_n - \theta = o_p(1)$ 이다. 따라서 $\sqrt{n}(X_n - \theta)^2 = o_p(1)$ 이다. 따라서

$$\sqrt{n}(g(X_n) - g(\theta)) = \sqrt{n}(X_n - \theta)\dot{g}(\theta) + o_p(1)O(1) \quad as \ X_n \rightarrow \theta$$

- $o_p(1)O(1) = o_p(1)$ 이므로

$$\sqrt{n}(g(X_n) - g(\theta)) = \sqrt{n}(X_n - \theta)\dot{g}(\theta) + o_p(1)$$

따라서

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{d} \sqrt{n}(X_n - \theta)\dot{g}(\theta)$$