수리통계학 2
6장. 확률변수의 극한

Seonghun Cho

Inha University
Department of Statistics
seonghun.cho@inha.ac.kr

2025 Fall

Outline

1. 확률변수의 극한
2. 대수의 법칙
3. 분포수렴
4. 중심극한정리

확률변수의 극한

- 확률변수의 수열 $\{X_n\}$에 대하여 X_n의 수렴에 대한 여러가지 정의

- **Almost sure convergence**:

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1$$

일 때, X_n이 X로 almost surely 수렴한다고 하며, 기호로 $X_n \xrightarrow{a.s.} X$와 같이 나타낸다.

- **확률수렴 (convergence in probability)**: 임의의 양수 $\epsilon > 0$에 대하여

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$$

일 때, X_n이 X로 확률수렴한다고 하며, 기호로 $X_n \xrightarrow{p} X$와 같이 나타낸다.

- **분포수렴 (convergence in distribution)**: 모든 실수 구간에서 X_n의 누적분포함수가 X의 누적분포함수로 수렴할 때, X_n이 X로 분포수렴한다고 하며, 기호로 $X_n \xrightarrow{d} X$와 같이 나타낸다.

대수의 법칙

- 표본의 크기 n이 커질 때 표본평균이 모평균으로 수렴한다는 것을 **대수의 법칙(Law of Large Numbers, LLN)**이라 한다.

Theorem (대수의 법칙)

$X_1, X_2, \ldots, X_n$이 기댓값 μ, 분산 σ^2인 확률분포부터의 랜덤샘플일 때 임의의 양수 $\epsilon > 0$에 대하여 다음이 성립한다.

$$\lim_{n\to\infty} P\left(|\overline{X}_n - \mu| > \epsilon\right) = 0.$$

즉, $\overline{X}_n \xrightarrow{p} \mu$이다.

Proof.

체비세프 부등식에 의하여 다음이 성립한다.

$$P\left(|\overline{X}_n - \mu| > \epsilon\right) \le \frac{\mathrm{Var}(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

n이 커질 때 $\sigma^2/(n\epsilon^2)$은 0으로 수렴하므로 $P\left(|\overline{X}_n - \mu| > \epsilon\right) \to 0$이다. □

분포수렴

Definition

$X_1, X_2, \ldots$이 누적분포함수 $F_1, F_2, \ldots$를 갖는 확률변수의 수열이고, 확률변수 X의 누적분포함수는 F라 하자. F가 연속인 모든 점 x에서

$$\lim_{n\to\infty} F_n(x) = F(x)$$

가 성립하면 X_n이 X로 **분포수렴(convergence in distribution)**한다고 한다. 이를 기호로 $X_n \xrightarrow{d} X$와 같이 나타낸다.

- 분포수렴에 대한 정의에서 한 가지 주의할 점은 모든 실수 x에서 $F_n(x)$이 $F(x)$로 수렴할 필요는 없고 F가 연속인 점 x에서만 $F_n(x) \to F(x)$이면 충분하다는 것이다.

여기 증명 이해하기

Example

$Z_1, Z_2, \dots, Z_n$이 $N(0,1)$로부터의 랜덤샘플일 때, 이들의 표본평균을 X_n이라고 하자. X_n은 정규분포 $N(0,1/n)$을 따르므로 X_n은 0으로 분포수렴한다고 볼 수 있다. (why?)

Solution

X_n의 누적분포함수의 극한은

$$F_n(x) = P(X_n \le x) = P(Z \le \sqrt{n}x) \longrightarrow \begin{cases} 0, & x < 0, \\ 0.5, & x = 0, \\ 1, & x > 0. \end{cases}$$

$X \equiv 0$의 누적분포함수는

$$F(x) = \begin{cases} 0, & x < 0, \\ 1, & x \ge 0. \end{cases}$$

$x = 0$을 제외한 모든 x에서 $\lim_{n\to\infty} F_n(x) = F(x)$이며, $x = 0$은 함수 $F(x)$의 불연속점이다. 따라서 $X_n \xrightarrow{d} 0$이다.

분포수렴

Example

$X_1, X_2, \ldots, X_n$이 $U(0, \theta)$로부터의 랜덤샘플일 때 $X_{(n)}$의 분포수렴 극한을 구하여라.

Solution

$X_{(n)} = \max(X_1, X_2, \ldots, X_n)$의 누적분포함수는

$$\begin{aligned} F_n(x) &= P(X_{(n)} \leq x) \\ &= P(X_1 \leq x) P(X_2 \leq x) \cdots P(X_n \leq x) \\ &= \left(\frac{x}{\theta}\right)^n, \quad 0 \leq x \leq \theta \end{aligned}$$

이므로 다음과 같다.

$$\lim_{n \to \infty} F_n(x) = \begin{cases} 0, & x < \theta, \\ 1, & x \geq \theta. \end{cases}$$

$X \equiv \theta$의 누적분포함수는 다음과 같다.

$$F(x) = \begin{cases} 0, & x < \theta, \\ 1, & x \geq \theta. \end{cases}$$

모든 x에 대하여 $\lim_{n \to \infty} F_n(x) = F(x)$이다. 따라서 $X_{(n)} \xrightarrow{d} \theta$이다.

분포수렴

- 앞의 예제는 $U(0, \theta)$로부터 생성된 샘플의 최댓값 $X_{(n)}$은 θ로 수렴한다는 사실을 보여준다.

- 다음 예제에서 우리는 $X_{(n)}$이 "얼마나 빨리" θ로 수렴하는지 살펴본다.

- $n(\theta - X_{(n)})$이 $\mathrm{Exp}(1/\theta)$로 수렴한다는 사실을 통해 $X_{(n)}$은 $1/n$의 속도로 θ로 수렴함을 보일 수 있다.

분포수렴

Example

$X_1, X_2 \dots, X_n$이 $U(0,\theta)$로부터의 랜덤샘플일 때 $Y_n = n(\theta - X_{(n)})$의 분포수렴 극한을 구하여라.

Solution

방향 맞춰주는게 좋다? 부.

그래서 1- 저렇게 함

Y_n의 누적분포함수는

$$\begin{aligned} F_n(y) &= P\left(n(\theta - X_{(n)}) \le y\right) \\ &= 1 - P\left(n(\theta - X_{(n)}) > y\right) \\ &= 1 - P\left(X_{(n)} < \theta - \frac{y}{n}\right) \\ &= 1 - \left(1 - \frac{y}{n\theta}\right)^n, \quad 0 < y < n\theta \\ &\longrightarrow 1 - e^{-y/\theta}, \quad y > 0 \end{aligned}$$

우리가 얘기할수 있

as $n \to \infty$. 이는 $Exp(1/\theta)$의 누적분포함수이므로 다음이 성립한다.

$$n(\theta - X_{(n)}) \xrightarrow{d} Exp(1/\theta).$$

분포수렴

- 작률생성함수를 이용하여 분포수렴을 보일 수도 있다.

Theorem

$X_1, X_2, \ldots$의 적률생성함수를 $M_1(t), M_2(t), \ldots$라 하고, X의 적률생성함수를 $M(t)$라 하자. 모든 실수 t에 대하여

$$\lim_{n\to\infty} M_n(t) = M(t)$$

가 성립하면 $X_n \xrightarrow{d} X$이다.

두 확률변수의 pdf는 달라도?
cdf는 같으면 같다
적률생성함수는 cdf의 또 다른 버전

분포수렴

Example

$X_n \sim \text{Poisson}(\lambda_n)$이며 $\lambda_n \to \infty$일 때,

$$Z_n = \frac{X_n - \mathbb{E}(X_n)}{\sqrt{\text{Var}(X_n)}} = \frac{X_n - \lambda_n}{\sqrt{\lambda_n}}$$

의 분포수렴 극한을 구하여라.

xn을 정규화한 새로운함수 zn은 어디로
가는가
zn의 분포가 어디로 분포수렴하겠나,
은 어려움 글서 mgf를 활용함
x는 이산확률변수라 계산하기위해 0부
터 무한대 e^tx~ 식

Solution

Z_n의 적률생성함수 $M_{Z_n}(t)$는

zn의

$$\begin{aligned}
M_{Z_n}(t) &= \mathbb{E}\left(e^{tZ_n}\right) = \mathbb{E}\left[\exp\left(\frac{t(X_n - \lambda_n)}{\sqrt{\lambda_n}}\right)\right] \\
&= \mathbb{E}\left[\exp\left(\frac{t}{\sqrt{\lambda_n}} X_n - t\sqrt{\lambda_n}\right)\right] \\
&= e^{-t\sqrt{\lambda_n}} \mathbb{E}\left[\exp\left(\frac{t}{\sqrt{\lambda_n}} X_n\right)\right] = e^{-t\sqrt{\lambda_n}} M_{X_n}(t/\sqrt{\lambda_n}) \\
&= e^{-t\sqrt{\lambda_n}} \exp\left(\lambda_n (e^{t/\sqrt{\lambda_n}} - 1)\right) = \exp\left(\lambda_n (e^{t/\sqrt{\lambda_n}} - 1) - t\sqrt{\lambda_n}\right)
\end{aligned}$$

분포수렴

Solution

(Continued)

$$\lim_{x\to 0}\frac{e^x-1-x}{x^2/2}=1$$

이므로 Z_n의 적률생성함수의 극한은 다음과 같다.

$$\begin{aligned}
M_{Z_n}(t) &= \exp\left(\lambda_n(e^{t/\sqrt{\lambda_n}}-1)-t\sqrt{\lambda_n}\right)\\
&= \exp\left(\frac{e^{t/\sqrt{\lambda_n}}-1-t/\sqrt{\lambda_n}}{t^2/(2\lambda_n)}\times\frac{t^2}{2}\right)\\
&\to e^{t^2/2} \quad as \quad n\to\infty.
\end{aligned}$$

$M(t)=e^{t^2/2}$*은 표준정규분포의 적률생성함수이므로* $Z_n \xrightarrow{d} N(0,1)$*이다.*

Example

$X \sim \text{Poisson}(900)$일 때 $P(X > 950)$의 근삿값은?

Solution

앞의 예제의 결과로부터 $\frac{X-900}{\sqrt{900}} \xrightarrow{d} N(0,1)$*이므로 다음과 같이 근사할 수 있다.*

$$\begin{aligned} P(X > 950) &= P\left(\frac{X-900}{\sqrt{900}} > \frac{950-900}{\sqrt{900}}\right) \\ &\approx P(N(0,1) > 1.67) = 0.04779. \end{aligned}$$

참고로 포아송분포로부터 직접 구한 $P(X > 950)$*의 값은 0.04712이다.*

중심극한정리

- 모집단의 확률분포에 상관없이 랜덤샘플의 표본평균을 표준화하면 자료의 수가 증가할 때 표본평균의 분포가 표준정규분포에 가까워진다는 것이 중심극한정리(Central Limit Theorem, CLT)이다.

Theorem (중심극한정리, CLT)

$X_1, X_2, \ldots, X_n$이 기댓값 μ, 분산 σ^2인 확률분포부터의 랜덤샘플이다. 그러면 다음이 성립한다.

$$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1).$$

이걸 보일 때 mgf를 사용
zn의 mgf를 구해서 mfg의 극한 값이 노말값의 그게 나온다를 보인다? 2페이지에 걸쳐서 나옴

중심극한정리

Proof.

Z_n의 적률생성함수 $M_{Z_n}(t)$는 다음과 같다.

$$\begin{aligned} M_{Z_n}(t) &= \mathbb{E}\left[\exp\left(\frac{t}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i-\mu}{\sigma}\right)\right] = \mathbb{E}\left[\exp\left(\frac{t}{\sqrt{n}}\sum_{i=1}^{n}Y_i\right)\right] \\ &= \left[\mathbb{E}\left\{\exp\left(\frac{t}{\sqrt{n}}Y\right)\right\}\right]^n = \left\{M_Y(t/\sqrt{n})\right\}^n \end{aligned}$$

위 식에서 $Y_i = \frac{X_i-\mu}{\sigma}$이고 M_Y는 Y_i의 적률생성함수이다. $M_Y(s)$를 0을 중심으로 테일러 전개하면 다음과 같다.

$$M_Y(s) = M_Y(0) + M_Y'(0)s + \frac{M_Y''(0)}{2}s^2 + \sum_{k=3}^{\infty}\frac{M_Y^{(k)}(0)}{k!}s^k.$$

$M_Y(0) = 1,\ M_Y'(0) = \mathbb{E}(Y) = 0,\ M_Y''(0) = \mathbb{E}(Y^2) = 1$이므로

$$M_Y(s) = 1 + \frac{s^2}{2} + \sum_{k=3}^{\infty}\frac{M_Y^{(k)}(0)}{k!}s^k$$

이다.

중심극한정리

Proof.

(Continued) 따라서

$$M_{Z_n}(t) = \left\{1 + \frac{t^2}{2n} + \sum_{k=3}^{\infty} \frac{M_Y^{(k)}(0)}{k!}\left(\frac{t}{\sqrt{n}}\right)^k\right\}^n$$
$$= \left[1 + \frac{1}{n}\left\{\frac{t^2}{2} + \sum_{k=3}^{\infty} \frac{M_Y^{(k)}(0)}{k!}\frac{t^k}{n^{k/2-1}}\right\}\right]^n$$

이다. 한편 $\lim_{n\to\infty} a_n = a$일 때 $\lim_{n\to\infty}(1 + a_n/n)^n = e^a$이므로 Z_n의 적률생성함수의 극한은 다음과 같다.

$$M_{Z_n}(t) \longrightarrow e^{t^2/2}$$

$e^{t^2/2}$은 표준정규분포의 적률생성함수이므로 $Z_n \xrightarrow{d} N(0,1)$이다. □

중심극한정리

- 중심극한정리는 서로 독립이며 같은 분포를 가지는 확률변수의 합은 정규분포에 가까이 간다는 것을 말해준다.

Example

$X_n \sim \text{Bin}(n, p)$이고 Y_i는 서로 독립이며 Bernoulli(p)를 따르는 확률변수일 때 $X_n = Y_1 + Y_2 + \cdots + Y_n$이므로 다음이 성립한다.

$$\frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

Example

$X_n \sim \text{Poisson}(\lambda_n)$이고 $Y_i \sim \text{Poisson}\,(\lambda_n/n)$이라 할 때 $X_n = Y_1 + Y_2 + \cdots + Y_n$라 나타낼 수 있다. 따라서 λ_n/n이 어느 정도 일정한 수준을 유지하면 중심극한정리에 의하여 $(X_n - \lambda_n)/\sqrt{\lambda_n}$은 표준정규분포에 분포수렴한다.

중심극한정리

Example

$X \sim \text{Bin}(100, 0.5)$일 때 $P(X \geq 60)$의 근삿값은?

Solution

X의 기댓값은 50이고 분산은 25이므로 $(X - 50)/\sqrt{25}$은 표준정규분포에 분포수렴한다. 따라서, 다음과 같이 근삿값을 구할 수 있다.

$$P(X \geq 60) = P\left(\frac{X - 50}{\sqrt{25}} \geq \frac{60 - 50}{\sqrt{25}}\right) \approx P(Z \geq 2) = 0.0228.$$