확률론 강의노트

MITZY

Last update : 2026.02

Contents

1 Logic

확률론은 "정의 $\to$ 정리 $\to$ 응용"의 형태로 전개되는 수학 과목이다. 따라서 강의노트의 첫 부분에서는, (1) 명제와 논리 기호, (2) 증명 방법의 기본 틀을 빠르게 정리한다.

1.1 명제와 논리 기호

수학에서 *명제*(statement, proposition)란 객관적으로 *참(True)* 또는 *거짓(False)*이 명확히 결정되는 문장(또는 식)을 말한다. 명제의 참/거짓을 *진리값*이라 한다.

두 명제 P, Q로부터 다음과 같은 합성 명제를 만들 수 있다.

- 부정 (negation): $\neg P$ 또는 $\sim P$는 "P가 아니다"를 뜻한다.
- 논리곱 (and): $P \wedge Q$는 "P이고 Q이다"를 뜻한다.
- 논리합 (or): $P \vee Q$는 "P이거나 Q이다"를 뜻한다.
- 함의 (implication): $P \Rightarrow Q$는 "만약 P이면 Q이다"를 뜻한다.
- 동치 (iff): $P \Leftrightarrow Q$는 "P일 필요충분조건으로 Q"를 뜻한다.

특히 함의 $P \Rightarrow Q$는 *P가 참이고 Q가 거짓일 때만 거짓*이며, P가 거짓인 경우에는 $P \Rightarrow Q$가 자동으로 참이 된다. (이 점이 처음에는 어색할 수 있으나, 반례를 통한 부정과 매우 잘 맞는다.)

1.2 필요조건과 충분조건

명제 $P \Rightarrow Q$에서

- Q는 P의 필요조건(necessary condition),
- P는 Q의 충분조건(sufficient condition)

이라 한다. 즉, "P이면 Q"에서 P가 성립하면 Q가 성립하므로 P는 Q를 보장하는 충분조건이다.

Example 1.1. "만약 $a \in \mathbb{N}$이면 $a \in \mathbb{R}$이다."는 참이다. 여기서 "$a \in \mathbb{R}$"은 "$a \in \mathbb{N}$"의 필요조건이지만, "$a \in \mathbb{R}$"만으로 "$a \in \mathbb{N}$"을 보장하지는 못하므로 충분조건은 아니다.

1.3 대표적인 증명 방식

수학적 증명에는 여러 방법이 있으나, 확률론 강의노트에서는 아래 네 가지가 특히 자주 사용된다.

(1) 직접 증명(direct proof): 가정에서 출발하여 논리적으로 결론을 유도한다.

(2) 대우 증명(contrapositive): $P \Rightarrow Q$ 대신 $\neg Q \Rightarrow \neg P$를 증명한다.

(3) 귀류법(proof by contradiction): 결론의 부정을 가정하여 모순을 유도한다.

(4) 수학적 귀납법(induction): 자연수에 대한 명제를 $P(1)$과 $P(k) \Rightarrow P(k+1)$로 증명한다.

1.4 구조화된 증명 예시 1: $\sqrt{2}$는 무리수

Theorem 1.2. $\sqrt{2} \notin \mathbb{Q}$.

구조화된 증명 가정**.** $\sqrt{2} \in \mathbb{Q}$라고 하자. 그러면 서로소인 정수 $n, m (m \neq 0)$가 존재하여 $\sqrt{2} = n/m$이다.
목표**.** 위 가정이 모순임을 보이자.

(1) 양변을 제곱하면 $2 = n^2/m^2$이므로 $n^2 = 2m^2$이다.

(2) n^2가 짝수이므로 n도 짝수이다. 따라서 어떤 정수 k가 존재하여 $n = 2k$이다.

(3) $n = 2k$를 $n^2 = 2m^2$에 대입하면 $4k^2 = 2m^2$, 즉 $m^2 = 2k^2$이다.

(4) m^2가 짝수이므로 m도 짝수이다.

(5) 따라서 n과 m은 공약수 2를 가지며, "서로소"라는 가정과 모순이다.

결론적으로 $\sqrt{2} \in \mathbb{Q}$라는 가정이 거짓이므로 $\sqrt{2} \notin \mathbb{Q}$이다. □

1.5 구조화된 증명 예시 2: 수학적 귀납법

Theorem 1.3. *모든* $n \in \mathbb{N}$*에 대하여* $\displaystyle\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$.

구조화된 증명 **목표.** 명제 $P(n) : \sum_{k=1}^{n} k = \frac{n(n+1)}{2}$가 모든 $n \in \mathbb{N}$에 대해 참임을 보이자.

(1) (기초 단계) $n = 1$이면 $\sum_{k=1}^{1} k = 1 = \frac{1 \cdot 2}{2}$이므로 $P(1)$은 참이다.

(2) (귀납 단계) 임의의 $k \in \mathbb{N}$에 대해 $P(k)$가 참이라고 가정하자:

$$\sum_{j=1}^{k} j = \frac{k(k+1)}{2}.$$

이제 $P(k+1)$을 보이자.

(3) 위 가정을 이용하면

$$\sum_{j=1}^{k+1} j = \left(\sum_{j=1}^{k} j\right) + (k+1) = \frac{k(k+1)}{2} + (k+1) = \frac{(k+1)(k+2)}{2}.$$

따라서 $P(k) \Rightarrow P(k+1)$이 성립한다.

(4) 기초 단계와 귀납 단계에 의해 모든 $n \in \mathbb{N}$에 대해 $P(n)$이 참이다.

□

1.6 양화사 기호와 자주 쓰는 약어

확률론에서는 "모든", "존재한다"와 같은 표현이 매우 빈번하므로 기호를 익혀두면 편하다.

- $\forall$: "모든" (for all, for every), "임의의".
- $\exists$: "존재한다" (there exists), "적어도 하나".
- $\exists!$: "유일하게 존재한다" (there exists uniquely).
- s.t.: such that, 조건을 서술할 때 사용.
- i.e.: id est, "즉".
- e.g.: for example, "예를 들어".

2 Basic Set theory

확률론은 결국 "집합 위에 정의된 측도"를 다루는 학문이므로, 집합론의 기초 표기부터 정리한다.

2.1 기본 표기

집합은 보통 대문자 $A, B, \ldots$ 로, 원소는 소문자 $a, b, \ldots$ 로 표기한다.

$$a \in A \quad (\text{ }a\text{는 }A\text{의 원소 }), \qquad a \notin A \quad (\text{ }a\text{는 }A\text{의 원소가 아님 }).$$

Definition 2.1 (전체집합과 여집합)**.** 어떤 "공간" (전체집합) U가 주어져 있고 $A \subseteq U$일 때,

$$A^c := U \setminus A$$

를 A의 *여집합* (complement)이라 한다.

Remark 2.2. 여집합 A^c는 반드시 *전체집합 U에 의존*한다. 확률론에서는 거의 항상 $U = \Omega$를 전체집합으로 두고 $A^c = \Omega \setminus A$로 쓴다.

원소가 없는 집합을 *공집합*이라 하고 $\emptyset$로 표기한다. 두 집합 A, B에 대해

$$A \subseteq B \iff (\forall a)(a \in A \Rightarrow a \in B)$$

이며, $A \subseteq B$와 $B \subseteq A$가 동시에 성립하면 $A = B$이다. 또한 $\emptyset \subseteq A$는 모든 집합 A에 대해 성립한다.

2.2 집합의 표현과 멱집합

집합은 원소를 나열하거나(유한 집합) 조건으로 정의할 수 있다.

$$A = \{1, 2\}, \qquad B = \{x \in \mathbb{N} : x \text{ is even}\}.$$

집합 A의 모든 부분집합들의 모임을 *멱집합* (power set)이라 하고 $\mathcal{P}(A)$ 또는 2^A로 쓴다.

2.3 집합 연산

두 집합 $A, B \subseteq U$에 대해 다음을 정의한다.

$$\begin{aligned}
A \cup B &:= \{x \in U : x \in A \text{ or } x \in B\} \quad (\text{합집합}),\\
A \cap B &:= \{x \in U : x \in A \text{ and } x \in B\} \quad (\text{교집합}),\\
A \setminus B &:= \{x \in U : x \in A \text{ and } x \notin B\} = A \cap B^c \quad (\text{차집합}),\\
A \triangle B &:= (A \setminus B) \cup (B \setminus A) \quad (\text{대칭차집합}).
\end{aligned}$$

2.4 드모르간 법칙과 분배법칙

아래 등식들은 확률론에서 사건 연산을 다룰 때 매번 사용된다.

Theorem 2.3 (기본 항등식)**.** *모든 $A, B, C \subseteq U$에 대해*

$$\begin{aligned}
(A \cup B)^c &= A^c \cap B^c, & (A \cap B)^c &= A^c \cup B^c,\\
A \cup (B \cap C) &= (A \cup B) \cap (A \cup C), & A \cap (B \cup C) &= (A \cap B) \cup (A \cap C).
\end{aligned}$$

구조화된 증명: 원소 추적(element-chasing). 첫 번째 등식 $(A \cup B)^c = A^c \cap B^c$만 보이면 나머지도 같은 방식으로 증명된다.

(1) 목표**.** 임의의 $x \in U$에 대해 $x \in (A \cup B)^c \iff x \in A^c \cap B^c$를 보이자.

(2) $x \in (A \cup B)^c$
$\iff x \notin (A \cup B)$ (여집합 정의)
$\iff (x \notin A)$ and $(x \notin B)$ (합집합 정의)
$\iff (x \in A^c)$ and $(x \in B^c)$ (여집합 정의)
$\iff x \in A^c \cap B^c$ (교집합 정의).

임의의 x에 대해 동치가 성립하므로 두 집합이 같다. □

2.5 지표가 있는 합집합/교집합

지표집합(index set) I와 집합족(family) $\{A_i\}_{i\in I}$가 주어졌을 때,

$$\bigcup_{i\in I} A_i := \{x : \exists i \in I \text{ s.t. } x \in A_i\}, \qquad \bigcap_{i\in I} A_i := \{x : \forall i \in I,\ x \in A_i\}.$$

특히 $I = \mathbb{N}$이면 *가산* 개의 합집합/교집합을 다룬다. 확률론에서 σ-algebra의 정의가 "가산 합집합"을 포함하는 이유가 바로 여기에 있다.

2.6 집합열의 극한: lim sup와 lim inf

집합열 $(A_n)_{n\in\mathbb{N}}$에 대해 다음을 정의한다.

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \{\omega : \omega \in A_n \text{ for infinitely many } n\},$$

$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \{\omega : \omega \in A_n \text{ for all but finitely many } n\}.$$

확률론에서는 $\limsup A_n$을 "A_n이 무한히 자주 일어난다" (i.o., infinitely often)라는 사건으로 해석한다.

Proposition 2.4. *항상* $\liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n$*가 성립한다. 또한*

$$\left(\limsup_{n\to\infty} A_n\right)^c = \liminf_{n\to\infty}(A_n^c), \qquad \left(\liminf_{n\to\infty} A_n\right)^c = \limsup_{n\to\infty}(A_n^c).$$

구조화된 증명 **(1)** $\liminf A_n \subseteq \limsup A_n$를 보이자. $\omega \in \liminf A_n$이면 어떤 n_0가 존재하여 모든 $k \geq n_0$에 대해 $\omega \in A_k$이다. 그러면 특히 무한히 많은 k에 대해 $\omega \in A_k$이므로 $\omega \in \limsup A_n$이다.

(2) 첫 번째 등식을 보이자.

$$(\limsup A_n)^c = \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right)^c = \bigcup_{n=1}^{\infty} \left(\bigcup_{k=n}^{\infty} A_k\right)^c = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c = \liminf A_n^c,$$

여기서 두 번째와 세 번째 등호는 드모르간 법칙(가산 버전)을 사용했다.

(3) 두 번째 등식도 같은 방식으로 증명된다.

□

2.7 데카르트 곱

두 집합 A, B의 *데카르트 곱*(Cartesian product)은

$$A \times B := \{(a, b) : a \in A,\ b \in B\}$$

로 정의한다. 예를 들어 $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$이며, 일반적으로 $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$로 정의한다.

3 Functions

집합과 함수는 확률론(더 넓게는 현대 수학)의 공통 언어이다. 확률변수(random variable)는 "확률공간 Ω에서 실수공간 $\mathbb{R}$로 가는 *가측 함수*"로 정의된다. 따라서 "함수"의 정의를 관계(relation) 관점에서 엄밀히 적어 두자.

3.1 함수의 정의

Definition 3.1 (함수)**.** 두 집합 A, B가 주어졌다고 하자. 부분집합 $G \subseteq A \times B$가 다음 두 조건을 만족하면, G를 A에서 B로 가는 *함수*(function)라 하고 $f : A \to B$로 표기한다.

(1) (존재) 모든 $a \in A$에 대해 어떤 $b \in B$가 존재하여 $(a, b) \in G$이다.

(2) (유일) $(a, b) \in G$이고 $(a, b') \in G$이면 $b = b'$이다.

이때 $(a, b) \in G$일 때 b를 $f(a)$로 쓰고 $f(a) = b$라 한다.

Remark 3.2. 조건 (존재)가 빠지면 "정의역의 일부에서만 값이 정해지는" *부분함수(partial function)*가 된다. 확률론에서는 보통 *정의역 전체*에서 정의된 함수를 다룬다.

정의역(domain)은 A, 공역(codomain)은 B이다. 치역(image)은

$$f(A) := \{f(a) : a \in A\} \subseteq B$$

이다.

3.2 단사/전사/전단사와 역함수

Definition 3.3 (단사, 전사, 전단사)**.** 함수 $f : A \to B$에 대해

- f가 *단사*(injective) $\iff$ $f(a) = f(a') \Rightarrow a = a'$.
- f가 *전사*(surjective) $\iff$ 모든 $b \in B$에 대해 어떤 $a \in A$가 존재하여 $f(a) = b$.
- f가 *전단사*(bijective) $\iff$ 단사이면서 전사.

Definition 3.4 (역함수)**.** $f : A \to B$가 전단사이면, 다음을 만족하는 함수 $f^{-1} : B \to A$가 유일하게 존재한다:

$$f^{-1}(b) = a \iff f(a) = b.$$

이 f^{-1}을 f의 *역함수*(inverse function)라 한다.

Theorem 3.5 (역함수의 존재 조건)**.** *함수 $f : A \to B$가 역함수를 가질 필요충분조건은 f가 전단사인 것이다.*

구조화된 증명 **(1) (⇒)** f가 역함수 $g : B \to A$를 가진다고 하자.

(1.1) *단사성:* $f(a) = f(a')$이면 양변에 g를 합성하여 $a = g(f(a)) = g(f(a')) = a'$.

(1.2) *전사성:* 임의의 $b \in B$에 대해 $b = f(g(b))$이므로 b는 f의 상이다.

(2) (⇐) f가 전단사라 하자. 임의의 $b \in B$에 대해 전사성으로 $f(a) = b$인 a가 존재한다. 단사성 때문에 그러한 a는 유일하다. 이 유일한 a를 $f^{-1}(b)$로 정의하면 $f^{-1} : B \to A$가 잘 정의되며 정의에 의해 f^{-1}은 f의 역함수이다.

□

3.3 역상(preimage)과 그 성질

확률론에서 가장 자주 쓰이는 연산은 *역상*이다.

Definition 3.6 (역상)**.** 함수 $f : A \to B$와 부분집합 $C \subseteq B$에 대해

$$f^{-1}(C) := \{a \in A : f(a) \in C\}$$

를 C의 *역상*(preimage)이라 한다.

Remark 3.7. $f^{-1}(C)$는 f가 전단사일 필요가 없다. 즉, "역상"은 언제나 정의되지만 "역함수"는 전단사일 때만 존재한다.

Proposition 3.8 (역상이 집합 연산을 보존)**.** *$f : A \to B$와 $C, D \subseteq B$에 대해*

(1) $f^{-1}(C \cup D) = f^{-1}(C) \cup f^{-1}(D)$.

(2) $f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$.

(3) $f^{-1}(C^c) = (f^{-1}(C))^c$ *(여기서 $C^c = B \setminus C$, $(f^{-1}(C))^c = A \setminus f^{-1}(C)$).*

(4) $f^{-1}\left(\bigcup_{n\in\mathbb{N}} C_n\right) = \bigcup_{n\in\mathbb{N}} f^{-1}(C_n)$, $\quad f^{-1}\left(\bigcap_{n\in\mathbb{N}} C_n\right) = \bigcap_{n\in\mathbb{N}} f^{-1}(C_n)$.

구조화된 증명(원소 추적). (1)만 보이면 나머지도 동일하다.

(1) 목표**.** 임의의 $a \in A$에 대해 $a \in f^{-1}(C \cup D) \iff a \in f^{-1}(C) \cup f^{-1}(D)$를 보이자.

(2) $a \in f^{-1}(C \cup D) \iff f(a) \in C \cup D \iff (f(a) \in C)$ or $(f(a) \in D) \iff (a \in f^{-1}(C))$ or $(a \in f^{-1}(D)) \iff a \in f^{-1}(C) \cup f^{-1}(D)$.

□

3.4 가산 집합과 비가산 집합

확률론에서는 "가산"과 "비가산"이 자주 등장한다. (예: 가산 합집합, 가산 생성계 등)

Definition 3.9 (가산 집합)**.** 집합 A가 *가산*(countable)이라 함은 A가 유한이거나, 또는 $\mathbb{N}$과 일대일 대응(전단사)인 경우를 말한다.

Theorem 3.10 (가산성의 기본 성질)**.** *다음이 성립한다.*

(1) *가산 집합의 부분집합은 가산이다.*

(2) *가산 집합 두 개의 합집합과 곱집합은 가산이다.*

(3) *가산 개의 가산 집합의 합집합은 가산이다.*

Remark 3.11. 대표적으로 $\mathbb{Z}, \mathbb{Q}$는 가산이고, $\mathbb{R}$은 비가산이다. 칸토어의 대각선 논법은 $(0, 1)$이 비가산임을 보이는데, 따라서 $\mathbb{R}$도 비가산이다.

4 Real number

확률론에서 실수공간 $\mathbb{R}$은 확률변수의 값공간이자, 르베그 측도와 적분이 정의되는 기본 무대이다. 여기서는 해석학적 배경 중 확률론에서 특히 중요한 두 가지를 강조한다:

- **(완비성)** 상한의 존재(least upper bound property),
- **(조밀성)** 유리수/무리수의 조밀성(density).

4.1 유리수와 실수

유리수는

$$\mathbb{Q} := \left\{ \frac{m}{n} : m, n \in \mathbb{Z},\ n \neq 0 \right\}$$

로 정의된다. 임의의 유리수는 분모/분자를 약분하여 서로소인 표현으로 만들 수 있으나, 정의 자체에 "서로소" 조건을 강제할 필요는 없다.

실수 $\mathbb{R}$은 엄밀하게는 "완비 순서체(complete ordered field)"로 공리화된다. 이 강의노트에서는 해석학에서처럼 $\mathbb{R}$이 그러한 성질을 만족한다고 *가정*하고 출발한다.

4.2 상계, 상한, 완비성

$A \subseteq \mathbb{R}$가 공집합이 아니라고 하자.

Definition 4.1 (위로 유계)**.** 실수 u가 존재하여 모든 $x \in A$에 대해 $x \leq u$이면 A를 *위로 유계*(bounded above)라 하고, 이러한 u를 A의 *상계*(upper bound)라 한다.

Definition 4.2 (상한)**.** A가 위로 유계일 때 실수 α가 다음을 만족하면 α를 A의 *상한*(supremum)이라 하고 $\alpha = \sup A$로 쓴다.

(1) α는 A의 상계이다: $\forall x \in A,\ x \leq \alpha$.

(2) α는 *가장 작은* 상계이다: 임의의 $\varepsilon > 0$에 대해 $\alpha - \varepsilon$는 상계가 아니다. 즉, $\exists x \in A$ s.t. $\alpha - \varepsilon < x$.

Theorem 4.3 (완비성 공리(least upper bound property))**.** *공집합이 아닌 $A \subseteq \mathbb{R}$가 위로 유계이면 $\sup A$가 존재하며 $\sup A \in \mathbb{R}$이다.*

Remark 4.4. 위 정리는 *실수의 정의에 포함되는 공리*로 이해하는 것이 가장 깔끔하다. 유리수체 $\mathbb{Q}$에서는 위로 유계인 집합이 상한을 가지더라도 그 상한이 $\mathbb{Q}$에 속하지 않을 수 있다. 예를 들어 $A = \{q \in \mathbb{Q} : q^2 < 2\}$는 $\mathbb{Q}$에서 위로 유계이지만 $\sup A = \sqrt{2} \notin \mathbb{Q}$이다.

4.3 아르키메데스 성질

Theorem 4.5 (아르키메데스 성질)**.** *자연수 집합 $\mathbb{N}$은 $\mathbb{R}$에서 위로 유계가 아니다. 즉, 임의의 $M \in \mathbb{R}$에 대해 $n > M$인 $n \in \mathbb{N}$이 존재한다.*

구조화된 증명(귀류법). 가정**.** $\mathbb{N}$이 위로 유계라고 가정하자.
목표**.** 모순을 도출하자.

(1) 가정에 의해 $\alpha := \sup \mathbb{N}$이 존재한다.

(2) $\varepsilon = 1$을 상한의 성질(2)에 대입하면, $\alpha - 1$은 상계가 아니다. 따라서 어떤 $n \in \mathbb{N}$가 존재하여 $\alpha - 1 < n$이다.

(3) 위 부등식에 1을 더하면 $\alpha < n + 1$이다. 그런데 $n + 1 \in \mathbb{N}$이므로 α가 $\mathbb{N}$의 상계라는 사실과 모순이다.

따라서 $\mathbb{N}$은 위로 유계가 아니다. □

4.4 정수 부분(floor/ceiling)의 존재

유리수의 조밀성을 증명할 때 "x보다 큰 가장 작은 정수"(ceiling) 같은 개념이 필요하다. 이를 엄밀히 다루려면 해석학의 표준 정리를 이용한다.

Lemma 4.6 (정수 부분의 존재)**.** *임의의 $x \in \mathbb{R}$에 대해 정수 $m \in \mathbb{Z}$가 존재하여*

$$m \leq x < m + 1$$

을 만족한다. (즉, $\lfloor x \rfloor$가 존재한다.)

구조화된 증명 목표**.** 어떤 정수 m이 $m \leq x < m + 1$을 만족함을 보이자.

(1) 아르키메데스 성질로부터 $N \in \mathbb{N}$가 존재하여 $N > |x|$이다. 그러면 특히 $-N < x < N$이다.

(2) 이제 $x + N > 0$이므로 집합
$$U := \{n \in \mathbb{N} : n > x + N\}$$
은 공집합이 아니다.

(3) 자연수의 well-ordering에 의해 U의 최소원 $n_0 := \min U$가 존재한다. 정의에 의해 $n_0 > x + N$이고, 최소성이므로 $n_0 - 1 \leq x + N$이다.

(4) $m := (n_0 - 1) - N \in \mathbb{Z}$로 두면,

$$m = (n_0 - 1) - N \leq x \quad \text{and} \quad x < n_0 - N = m + 1$$

이므로 $m \leq x < m + 1$을 얻는다.

□

Remark 4.7. 위 증명은 (i) 아르키메데스 성질과 (ii) 자연수의 well-ordering 원리를 조합한 전형적인 방법이다. 동등한 결과로서 x 이하의 정수 집합의 상한을 취하는 방식의 증명도 널리 쓰인다.

4.5 유리수의 조밀성

Theorem 4.8 (유리수의 조밀성)**.** *임의의 $a < b$에 대해 $a < q < b$인 유리수 $q \in \mathbb{Q}$가 존재한다.*

구조화된 증명 **(1)** $b - a > 0$이므로 아르키메데스 성질에 의해 $\frac{1}{n} < b - a$를 만족하는 $n \in \mathbb{N}$이 존재한다.

(2) Lemma 4.6를 $x = na$에 적용하여 $m \leq na < m + 1$인 정수 m을 잡자. 그러면 $na < m + 1$이고, 한편 $m + 1 \leq na + 1 < na + n(b - a) = nb$이다. 따라서

$$na < m + 1 < nb.$$

(3) $q := \frac{m+1}{n}$으로 두면 $q \in \mathbb{Q}$이며 위 부등식을 n으로 나누어 $a < q < b$를 얻는다.

□

Remark 4.9. 같은 방식으로 무리수도 조밀함을 보일 수 있다. 예를 들어 (a, b) 안의 유리수 q를 하나 고른 뒤, 충분히 작은 양수 δ에 대해 $q + \delta\sqrt{2} \in (a, b)$가 되도록 만들면 된다.

5 Sequence

확률론에서는 확률변수열 (X_n)의 수렴을 다루기 전에, 실수열 (a_n)의 수렴을 먼저 이해하는 것이 중요하다. (실수열은 "정의역이 $\mathbb{N}$인 함수"라는 관점에서 확률변수열과 구조가 동일하다.)

5.1 수열의 정의

Definition 5.1 (수열)**.** 수열 $(a_n)_{n\in\mathbb{N}}$은 정의역이 $\mathbb{N}$이고 공역이 $\mathbb{R}$인 함수 $a : \mathbb{N} \to \mathbb{R}$를 말한다. 관례적으로 $a(n) = a_n$이라 쓴다.

예: $(a_n) = (1, 1/2, 1/3, \ldots)$는 $a_n = 1/n$으로 정의된다.

5.2 수렴

Definition 5.2 (수열의 수렴)**.** 실수열 (a_n)과 실수 x에 대해, 임의의 $\varepsilon > 0$에 대하여 어떤 $N \in \mathbb{N}$이 존재하여 모든 $n \geq N$에 대해

$$|a_n - x| < \varepsilon$$

이면 $a_n \to x$라고 하고 $\lim_{n\to\infty} a_n = x$로 쓴다.

Theorem 5.3 (극한의 유일성)**.** *실수열 (a_n)이 x로 수렴하고 y로도 수렴하면 $x = y$이다.*

구조화된 증명 **(1)** $a_n \to x$이므로 $\varepsilon/2$에 대해 어떤 N_1이 존재하여 $n \geq N_1$이면 $|a_n - x| < \varepsilon/2$.

(2) $a_n \to y$이므로 $\varepsilon/2$에 대해 어떤 N_2가 존재하여 $n \geq N_2$이면 $|a_n - y| < \varepsilon/2$.

(3) $n \geq N := \max\{N_1, N_2\}$이면 삼각부등식으로

$$|x - y| \leq |x - a_n| + |a_n - y| < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

(4) 임의의 $\varepsilon > 0$에 대해 $|x - y| < \varepsilon$이므로 $|x - y| = 0$, 즉 $x = y$.

□

5.3 유계성과 코시 수열

Definition 5.4 (유계 수열)**.** 실수열 (a_n)이 *유계*(bounded)라 함은 어떤 $M > 0$가 존재하여 모든 n에 대해 $|a_n| \le M$인 것이다.

Theorem 5.5 (수렴 $\Rightarrow$ 유계)**.** *수렴하는 실수열은 유계이다.*

구조화된 증명 가정**.** $a_n \to x$라고 하자.

(1) $\varepsilon = 1$에 대해 어떤 N이 존재하여 $n \ge N$이면 $|a_n - x| < 1$이다.

(2) 따라서 $n \ge N$이면 $|a_n| \le |x| + 1$이다.

(3) 유한 집합 $\{|a_1|, \dots, |a_{N-1}|\}$의 최댓값을 M_0라 두고, $M := \max\{M_0, |x| + 1\}$로 두면 모든 n에 대해 $|a_n| \le M$이다.

□

Remark 5.6. 유계 수열이 항상 수렴하는 것은 아니다. 예: $a_n = (-1)^n$은 유계이지만 수렴하지 않는다.

Definition 5.7 (코시 수열)**.** 실수열 (a_n)이 *코시*(Cauchy)라 함은, 임의의 $\varepsilon > 0$에 대해 어떤 N이 존재하여 $m, n \ge N$이면

$$|a_m - a_n| < \varepsilon$$

인 것이다.

Theorem 5.8 (실수의 완비성과 코시 수열)**.** *실수열* (a_n)*이 수렴할 필요충분조건은 코시 수열인 것이다.*

Remark 5.9. 위 정리는 실수의 완비성(completeness)을 "수열" 언어로 표현한 것이다. $\mathbb{Q}$에서는 코시 수열이지만 $\mathbb{Q}$에서 수렴하지 않는 수열이 존재한다(예: $\sqrt{2}$로 수렴하는 유리수열).

5.4 부분수열과 Bolzano–Weierstrass

Definition 5.10 (부분수열)**.** (a_n)의 부분수열(subsequence)은 증가하는 자연수열 $n_k \uparrow \infty$를 택하여 (a_{n_k})로 만든 수열이다.

Theorem 5.11 (Bolzano–Weierstrass)**.** *유계인 실수열은 수렴하는 부분수열을 가진다.*

Remark 5.12. 이 정리는 확률론에서 "부분수열 추출"의 원형이다. 예: 확률수렴 $X_n \xrightarrow{P} X$로부터 a.s. 수렴 부분수열을 뽑는 데에도 유사한 아이디어가 사용된다.

5.5 급수(Series) 간단 메모

수열 (a_n)의 부분합 $S_N = \sum_{n=1}^{N} a_n$이 $N \to \infty$일 때 수렴하면 급수 $\sum_{n=1}^{\infty} a_n$가 수렴한다고 한다. 절대수렴 $\sum |a_n| < \infty$이면 급수는 수렴한다.

6 Probability space

확률론의 현대적(측도론적) 정의는 다음 세 가지 데이터로 이루어진다:

$$(\Omega, \mathcal{F}, \mathbb{P}).$$

여기서 Ω는 표본공간(sample space), $\mathcal{F}$는 사건(event)들의 모임, $\mathbb{P}$는 확률(probability measure)이다. 핵심은 "*Ω의 모든 부분집합에 확률을 줄 수는 없다*"는 사실이며, 이를 해결하기 위해 $\mathcal{F}$를 특별한 집합족(= σ-algebra)으로 제한한다.

6.1 σ-algebra

Definition 6.1 (σ-algebra)**.** 집합 Ω의 멱집합 2^{Ω}의 부분집합 $\mathcal{F} \subseteq 2^{\Omega}$가 다음을 만족하면 $\mathcal{F}$를 σ-algebra 라 한다.

(1) $\Omega \in \mathcal{F}$.

(2) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F} \quad$ (여기서 $A^c = \Omega \setminus A$).

(3) $A_1, A_2, \cdots \in \mathcal{F} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.

$(\Omega, \mathcal{F})$를 *가측공간*(measurable space)이라 하며, $\mathcal{F}$의 원소 A를 확률론에서는 *사건*(event), 측도론에서는 *가측집합*(measurable set)이라 부른다.

Theorem 6.2 (기본 성질)**.** *$(\Omega, \mathcal{F})$가 가측공간이면 다음이 성립한다.*

(1) $\emptyset \in \mathcal{F}$.

(2) $A_1, A_2, \cdots \in \mathcal{F} \Rightarrow \bigcap_{n\in\mathbb{N}} A_n \in \mathcal{F}$.

구조화된 증명 **(1)** $\emptyset = \Omega^c$이므로 (2)에 의해 $\emptyset \in \mathcal{F}$.

(2) 드모르간 법칙을 사용하면

$$\bigcap_{n\in\mathbb{N}} A_n = \left(\bigcup_{n\in\mathbb{N}} A_n^c\right)^c.$$

각 $A_n \in \mathcal{F}$이므로 $A_n^c \in \mathcal{F}$이고, (3)에 의해 $\bigcup_n A_n^c \in \mathcal{F}$, 마지막으로 여집합을 취해도 $\mathcal{F}$에 남으므로 $\bigcap_n A_n \in \mathcal{F}$.

□

6.2 생성된 σ-algebra

Definition 6.3 ($\sigma(\mathcal{A})$)**.** 집합족 $\mathcal{A} \subseteq 2^{\Omega}$가 주어졌을 때, $\mathcal{A}$를 포함하는 σ-algebra들 중 가장 작은 것을 $\sigma(\mathcal{A})$라 한다. 즉

$$\sigma(\mathcal{A}) := \bigcap\{\mathcal{G} : \mathcal{G} \text{ is a } \sigma\text{-algebra on } \Omega \text{ and } \mathcal{A} \subseteq \mathcal{G}\}.$$

Remark 6.4. 단일 집합 $A \subseteq \Omega$에 대해 $\sigma(A)$라고 쓰는 것은 $\sigma(\{A\})$를 의미한다.

6.3 π-system과 λ-system, Dynkin의 π-λ 정리

Definition 6.5 (π-system)**.** 집합족 $\mathcal{P} \subseteq 2^{\Omega}$가 교집합에 대해 닫혀 있으면, 즉 $A, B \in \mathcal{P} \Rightarrow A \cap B \in \mathcal{P}$ 이면 $\mathcal{P}$를 π-system이라 한다.

Definition 6.6 (λ-system (Dynkin system))**.** 집합족 $\mathcal{L} \subseteq 2^{\Omega}$가 다음을 만족하면 λ-system이라 한다.

(1) $\Omega \in \mathcal{L}$.

(2) $A \in \mathcal{L} \Rightarrow A^c \in \mathcal{L}$.

(3) 서로소인 $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{L}$에 대해 $\bigcup_n A_n \in \mathcal{L}$.

Lemma 6.7 (λ-system의 유도 성질)**.** *$\mathcal{L}$이 λ-system이면 다음이 성립한다.*

(1) *$A, B \in \mathcal{L}$이고 $A \subseteq B$이면 $B \setminus A \in \mathcal{L}$.*

(2) *$A_1 \subseteq A_2 \subseteq \cdots$인 증가열 $(A_n) \subseteq \mathcal{L}$에 대해 $\bigcup_n A_n \in \mathcal{L}$.*

구조화된 증명 **(1)** (1) $A \subseteq B$이면 A와 B^c는 서로소이다. 따라서 λ-system 성질로 $B^c \cup A \in \mathcal{L}$이고, 다시 여집합을 취해

$$(B^c \cup A)^c = B \cap A^c = B \setminus A \in \mathcal{L}.$$

(2) (2) $B_1 := A_1$, $B_n := A_n \setminus A_{n-1}$ ($n \geq 2$)로 두면 (B_n)은 서로소이고 (1)에 의해 각 $B_n \in \mathcal{L}$이다. 또한 $\bigcup_n A_n = \bigcup_n B_n$이므로 λ-system 성질로 $\bigcup_n A_n \in \mathcal{L}$.

□

Theorem 6.8 (Dynkin의 π-λ 정리)**.** *$\mathcal{P}$가 π-system이고, $\mathcal{L}$이 $\mathcal{P}$를 포함하는 λ-system이면*

$$\sigma(\mathcal{P}) \subseteq \mathcal{L}.$$

구조화된 증명 전략**.** $\mathcal{P}$를 포함하는 가장 작은 λ-system을 먼저 만들고, $\mathcal{P}$가 π-system이면 그 집합족이 사실상 σ-algebra임을 보인다.

(1) $\lambda(\mathcal{P})$를 $\mathcal{P}$를 포함하는 모든 λ-system들의 교집합으로 정의하자. 그러면 $\lambda(\mathcal{P})$는 $\mathcal{P}$를 포함하는 가장 작은 λ-system이다.

(2) $\mathcal{L}$이 $\mathcal{P}$를 포함하는 λ-system이므로 정의에 의해 $\lambda(\mathcal{P}) \subseteq \mathcal{L}$이다. 따라서 목표 $\sigma(\mathcal{P}) \subseteq \mathcal{L}$를 위해서는 $\sigma(\mathcal{P}) \subseteq \lambda(\mathcal{P})$만 보이면 충분하다.

(3) 이제 $D := \lambda(\mathcal{P})$라 두고, D가 σ-algebra임을 보이자. 이를 위해 먼저 D가 *유한 교집합*에 대해 닫혀 있음을 보인다.

(3.1) 임의로 $A \in \mathcal{P}$를 고정하고

$$D_A := \{B \subseteq \Omega : B \in D \text{ and } A \cap B \in D\}$$

라 두자. D_A가 λ-system임을 보이자.

(3.1.1) $\Omega \in D$이고 $A \cap \Omega = A \in \mathcal{P} \subseteq D$이므로 $\Omega \in D_A$.

(3.1.2) $B \in D_A$이면 $B \in D$이고 $A \cap B \in D$이다. λ-system 성질로 $B^c \in D$. 또한 $A \cap B^c = A \setminus (A \cap B)$이고 $A \cap B \subseteq A$이므로 λ-system의 차집합 닫힘성(유도 성질)으로 $A \cap B^c \in D$. 따라서 $B^c \in D_A$.

(3.1.3) 서로소인 $(B_n) \subseteq D_A$에 대해 $B_n \in D$이며 $A \cap B_n \in D$이다. 또한 $(A \cap B_n)$도 서로소이므로 λ-system 성질로

$$\bigcup_n B_n \in D, \qquad A \cap \Big(\bigcup_n B_n\Big) = \bigcup_n (A \cap B_n) \in D.$$

따라서 $\bigcup_n B_n \in D_A$.

(3.2) $\mathcal{P} \subseteq D_A$임을 보이자. 임의의 $C \in \mathcal{P}$에 대해, $\mathcal{P}$가 π-system이므로 $A \cap C \in \mathcal{P} \subseteq D$이고 또한 $C \in \mathcal{P} \subseteq D$이므로 $C \in D_A$.

(3.3) D는 $\mathcal{P}$를 포함하는 가장 작은 λ-system이므로 $D \subseteq D_A$이다. 따라서 모든 $B \in D$에 대해 $A \cap B \in D$가 성립한다. 즉, *$\mathcal{P}$의 원소와의 교집합에 대해 D는 닫혀 있다.*

(3.4) 이제

$$C := \{A \in D : \forall B \in D,\ A \cap B \in D\}$$

라 두자. (3.3)에 의해 $\mathcal{P} \subseteq C$임을 알고 있다. C가 λ-system임을 보이면 $D \subseteq C$가 되어 D가 유한 교집합에 대해 닫혀 있음을 얻는다.

(3.4.1) $\Omega \in D$이고 임의의 $B \in D$에 대해 $\Omega \cap B = B \in D$이므로 $\Omega \in C$.

(3.4.2) $A \in C$이면 $A \in D$이므로 $A^c \in D$. 또한 임의의 $B \in D$에 대해 $A^c \cap B = B \setminus (A \cap B)$이고 $A \cap B \in D$ (정의)이며 $A \cap B \subseteq B$이므로 D의 차집합 닫힘성으로 $A^c \cap B \in D$. 따라서 $A^c \in C$.

(3.4.3) 서로소인 $(A_n) \subseteq C$에 대해 임의의 $B \in D$를 고정하자. 그러면 $(A_n \cap B)$도 서로소이고 각 $A_n \cap B \in D$이므로

$$\Big(\bigcup_n A_n\Big) \cap B = \bigcup_n (A_n \cap B) \in D.$$

또한 $\bigcup_n A_n \in D$이므로 $\bigcup_n A_n \in C$.

(3.5) 결론적으로 C는 λ-system이고 $\mathcal{P} \subseteq C \subseteq D$이므로 최소성에 의해 $D \subseteq C$. 따라서 임의의 $A, B \in D$에 대해 $A \cap B \in D$이다.

(4) 유한 교집합 닫힘성을 이용하여 D가 σ-algebra임을 보이자.

(4.1) D는 λ-system이므로 $\Omega \in D$, 여집합에 대해 닫혀 있다.

(4.2) 또한 D는 유한 교집합에 대해 닫혀 있으므로 유한 합집합에도 닫혀 있다:

$$A \cup B = (A^c \cap B^c)^c.$$

(4.3) 임의의 가산 집합족 $(A_n) \subseteq D$에 대해

$$B_1 := A_1, \qquad B_n := A_n \setminus \bigcup_{k<n} A_k \quad (n \geq 2)$$

로 두면 (B_n)은 서로소이며 각 $B_n \in D$이다(차집합은 (3.1.2) 과정에서 보인 유도 성질로 D에 남는다). 그리고 $\bigcup_n A_n = \bigcup_n B_n$이므로 λ-system의 서로소 가산 합집합 닫힘성으로 $\bigcup_n A_n \in D$.

(5) 따라서 $D = \lambda(\mathcal{P})$는 σ-algebra이고 $\mathcal{P} \subseteq D$이므로 $\sigma(\mathcal{P}) \subseteq D$이다.

(6) 반대로 $\sigma(\mathcal{P})$는 σ-algebra이므로 특히 λ-system이며 $\mathcal{P}$를 포함한다. 따라서 최소성으로 $D = \lambda(\mathcal{P}) \subseteq \sigma(\mathcal{P})$이다.

(7) 결론적으로 $D = \sigma(\mathcal{P})$이고, (2)에서 $\lambda(\mathcal{P}) \subseteq \mathcal{L}$였으므로 $\sigma(\mathcal{P}) \subseteq \mathcal{L}$이다.

□

Corollary 6.9 (측도의 유일성(자주 쓰는 형태)). *두 확률측도 $\mathbb{P}_1, \mathbb{P}_2$가 어떤 π-system $\mathcal{P}$ 위에서 일치하면 (즉, 모든 $A \in \mathcal{P}$에 대해 $\mathbb{P}_1(A) = \mathbb{P}_2(A)$), $\sigma(\mathcal{P})$ 위에서도 일치한다.*

구조화된 증명 가정**.** $\mathbb{P}_1, \mathbb{P}_2$는 $\mathcal{P}$ 위에서 일치한다.
목표**.** 모든 $A \in \sigma(\mathcal{P})$에 대해 $\mathbb{P}_1(A) = \mathbb{P}_2(A)$임을 보이자.

(1) $\mathcal{L} := \{A \in \sigma(\mathcal{P}) : \mathbb{P}_1(A) = \mathbb{P}_2(A)\}$라 두자.

(2) $\mathcal{L}$이 λ-system임을 보이자.

(2.1) $\Omega \in \mathcal{L}$이다. (두 확률측도는 $\mathbb{P}_i(\Omega) = 1$)

(2.2) $A \in \mathcal{L}$이면 $A^c \in \mathcal{L}$이다. (확률의 여집합 성질)

(2.3) 서로소인 $(A_n) \subseteq \mathcal{L}$에 대해

$$\mathbb{P}_1\Big(\bigcup_n A_n\Big) = \sum_n \mathbb{P}_1(A_n) = \sum_n \mathbb{P}_2(A_n) = \mathbb{P}_2\Big(\bigcup_n A_n\Big)$$

이므로 $\bigcup_n A_n \in \mathcal{L}$이다.

(3) 가정에 의해 $\mathcal{P} \subseteq \mathcal{L}$이다.

(4) Dynkin의 π-λ 정리에 의해 $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.

(5) 따라서 모든 $A \in \sigma(\mathcal{P})$에 대해 $\mathbb{P}_1(A) = \mathbb{P}_2(A)$이다.

□

Remark 6.10. 이 결과 덕분에 "모든 Borel 집합"에서 등식을 확인하지 않고, 생성계(generator) 위에서만 확인하여 분포(측도)의 일치성을 증명할 수 있다.

6.4 확률측도

Definition 6.11 (확률측도)**.** 다음 성질을 만족하는 함수 $\mathbb{P} : \mathcal{F} \to [0, 1]$를 확률(확률측도)라 한다.

(1) $\mathbb{P}(\emptyset) = 0$.

(2) $\mathbb{P}(\Omega) = 1$.

(3) 서로소인 $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F}$에 대해

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n).$$

(3)을 *가산 가법성*(σ-additivity)이라 한다.

Definition 6.12 (확률공간)**.** $(\Omega, \mathcal{F}, \mathbb{P})$를 *확률공간*(probability space)이라 한다.

6.5 확률의 기본 성질

Theorem 6.13. *모든 $A, B, A_n \in \mathcal{F}$에 대해 다음이 성립한다.*

(1) **(**단조성**)** $A \subseteq B$*이면* $\mathbb{P}(A) \le \mathbb{P}(B)$.

(2) **(**가산 아랫가법성**; Boole)** $\mathbb{P}(\bigcup_n A_n) \le \sum_n \mathbb{P}(A_n)$.

(3) **(**증가열 연속성**)** $A_1 \subseteq A_2 \subseteq \cdots$ *이면* $\mathbb{P}(\bigcup_n A_n) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

(4) **(**감소열 연속성**)** $A_1 \supseteq A_2 \supseteq \cdots$ *이면* $\mathbb{P}(\bigcap_n A_n) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

구조화된 증명(핵심 아이디어만). **(1)** (1) $B = A \cup (B \setminus A)$이고 A와 $B \setminus A$는 서로소이므로

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \ge \mathbb{P}(A).$$

(2) (2) $B_1 = A_1$, $B_n = A_n \setminus \bigcup_{k<n} A_k$로 두면 (B_n)는 서로소이고 $\bigcup_n B_n = \bigcup_n A_n$이다. 따라서

$$\mathbb{P}\Big(\bigcup_n A_n\Big) = \sum_n \mathbb{P}(B_n) \le \sum_n \mathbb{P}(A_n).$$

(3) (3) $B_1 = A_1$, $B_n = A_n \setminus A_{n-1}$로 두면 (B_n)는 서로소이고 $\bigcup_{k=1}^n B_k = A_n$, 또한 $\bigcup_n A_n = \bigcup_n B_n$이다. 가산 가법성을 적용하면

$$\mathbb{P}\Big(\bigcup_n A_n\Big) = \sum_n \mathbb{P}(B_n) = \lim_{n\to\infty} \sum_{k=1}^n \mathbb{P}(B_k) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(4) (4) (3)을 여집합에 적용하면 된다: $A_n \downarrow A$이면 $A_n^c \uparrow A^c$이므로 (3)으로부터 $\mathbb{P}(A^c) = \lim \mathbb{P}(A_n^c) = \lim(1 - \mathbb{P}(A_n))$, 따라서 $\mathbb{P}(A) = \lim \mathbb{P}(A_n)$.

□

6.6 영확률 집합과 a.s.

Definition 6.14 (null set / negligible set)**.** $A \in \mathcal{F}$이고 $\mathbb{P}(A) = 0$이면 A를 *영확률 집합*(null set)이라 한다.
또한 $N \subseteq A$인 임의의 부분집합 N을 *무시 가능한 집합*(negligible set)이라 부른다. 이때 N은 *가측일 필요가 없다.* (측도의 완비성과 관련)

Theorem 6.15. *가산 개의 negligible set의 합집합은 negligible set이다.*

구조화된 증명 **(1)** 각 N_k가 어떤 null set $A_k \in \mathcal{F}$의 부분집합이라고 하자: $N_k \subseteq A_k$, $\mathbb{P}(A_k) = 0$.

(2) 그러면 $\bigcup_k N_k \subseteq \bigcup_k A_k$이고, Theorem 6.13(2)로부터

$$\mathbb{P}\Big(\bigcup_k A_k\Big) \le \sum_k \mathbb{P}(A_k) = 0.$$

(3) 따라서 $\bigcup_k A_k$는 null set이고, $\bigcup_k N_k$는 그 부분집합이므로 negligible set이다.

□

Definition 6.16 (almost surely)**.** 어떤 성질 $P(\omega)$가

$$\mathbb{P}(\{\omega \in \Omega : P(\omega) \text{ 가 성립}\}) = 1$$

을 만족하면 "P가 거의 확실하게(almost surely, a.s.) 성립한다"고 한다. 동치로 $\mathbb{P}(\{\omega : P(\omega) \text{ 불성립}\}) = 0$이다.

7 Topological properties of Real number

확률변수는 Ω에서 $\mathbb{R}$로 가는 *가측 함수*이므로, $\mathbb{R}$ 위의 기본적인 위상/가측 구조(열린집합, Borel σ-algebra)를 정리한다.

7.1 거리공간과 열린집합

Definition 7.1 (거리공간)**.** 집합 X와 함수 $d : X \times X \to [0, \infty)$가 다음을 만족하면 (X, d)를 *거리공간*(metric space)이라 한다.

(1) $d(x, y) = 0 \iff x = y$,

(2) $d(x, y) = d(y, x)$,

(3) $d(x, z) \le d(x, y) + d(y, z)$.

실수공간에서는 $d(x, y) = |x - y|$를 표준 거리로 쓴다.

Definition 7.2 (근방과 열린집합)**.** $x \in \mathbb{R}$, $r > 0$에 대해

$$B(x, r) := \{y \in \mathbb{R} : |x - y| < r\} = (x - r, x + r)$$

를 *열린 공(ball)* 또는 x의 r-근방이라 한다.

집합 $U \subseteq \mathbb{R}$가 모든 $x \in U$에 대해 어떤 $r > 0$가 존재하여 $B(x, r) \subseteq U$이면 U를 *열린집합*(open set)이라 한다.

Theorem 7.3 (열린집합의 안정성)**.** *임의의 열린집합들의 합집합은 열린집합이고, 유한 개 열린집합들의 교집합은 열린집합이다.*

7.2 Borel σ-algebra

Definition 7.4 (Borel σ-algebra)**.** $\mathbb{R}$의 모든 열린집합들의 모임을

$$\mathcal{O}(\mathbb{R}) := \{U \subseteq \mathbb{R} : U \text{ is open}\}$$

라 하자. 이때

$$\mathcal{B}(\mathbb{R}) := \sigma(\mathcal{O}(\mathbb{R}))$$

를 *Borel σ-algebra*라 하고, 원소를 *Borel 집합*이라 한다.

Remark 7.5. $\mathcal{B}(\mathbb{R})$는 "열린집합을 포함하고, 여집합과 가산 합집합에 대해 닫혀 있는 가장 작은 집합족"이다. 따라서 열린집합은 물론 닫힌집합도 Borel 집합이고, 가산 개의 열린/닫힌집합에 대한 합집합/교집합으로 만들어지는 집합들은 모두 Borel 집합이다.

7.3 생성계: 구간만으로도 충분

Borel σ-algebra는 다양한 간단한 생성계로 표현할 수 있다. 중요한 점은 *"하나의 구간" 이 아니라 "구간들의 모임"*이 생성계라는 것이다.

Theorem 7.6 (Borel σ-algebra의 표준 생성계)**.** *다음 집합족들은 모두 $\mathcal{B}(\mathbb{R})$를 생성한다.*

(1) $\mathcal{I}_o := \{(a,b) : a < b,\ a,b \in \mathbb{R}\}$ *(열린 유계구간 전체),*

(2) $\mathcal{I}_{o,\mathbb{Q}} := \{(a,b) : a < b,\ a,b \in \mathbb{Q}\}$ *(유리수 끝점만 사용),*

(3) $\mathcal{H} := \{(-\infty, b) : b \in \mathbb{R}\}$, $\mathcal{H}_c := \{(-\infty, b] : b \in \mathbb{R}\}$.

즉,

$$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{I}_o) = \sigma(\mathcal{I}_{o,\mathbb{Q}}) = \sigma(\mathcal{H}) = \sigma(\mathcal{H}_c).$$

구조화된 증명(대표적인 포함관계). 핵심은 "열린집합은 (서로소일 필요 없는) 가산 개의 열린구간의 합집합" 이라는 사실이다. 여기서는 $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{I}_{o,\mathbb{Q}})$만 보인다. 나머지도 같은 아이디어로 처리된다.

(1) $\mathcal{I}_{o,\mathbb{Q}} \subseteq \mathcal{O}(\mathbb{R})$이므로 $\sigma(\mathcal{I}_{o,\mathbb{Q}}) \subseteq \sigma(\mathcal{O}(\mathbb{R})) = \mathcal{B}(\mathbb{R})$.

(2) 반대로, 임의의 열린집합 $U \in \mathcal{O}(\mathbb{R})$를 잡자. 각 $x \in U$에 대해 어떤 $r_x > 0$가 존재하여 $(x - r_x, x + r_x) \subseteq U$.

(3) 유리수는 조밀하므로 각 $x \in U$에 대해 $a_x, b_x \in \mathbb{Q}$를 $x \in (a_x, b_x) \subseteq (x - r_x, x + r_x) \subseteq U$가 되도록 고를 수 있다.

(4) 따라서 U는 유리수 끝점을 갖는 열린구간들의 합집합으로 표현된다:

$$U = \bigcup_{x \in U} (a_x, b_x).$$

(5) 하지만 $\mathbb{Q} \times \mathbb{Q}$는 가산이므로, 서로 다른 구간들의 집합 $\{(a_x, b_x) : x \in U\}$은 가산 개 이하이다. 따라서 U는 *가산* 합집합으로 쓸 수 있고, 곧 $U \in \sigma(\mathcal{I}_{o,\mathbb{Q}})$이다.

따라서 $\mathcal{O}(\mathbb{R}) \subseteq \sigma(\mathcal{I}_{o,\mathbb{Q}})$이고, 양변에 $\sigma(\cdot)$를 취하면 $\mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{I}_{o,\mathbb{Q}})$이다. □

Remark 7.7. Theorem 7.6는 "확률변수의 가측성" 을 확인할 때 검사해야 하는 집합들을 크게 줄여준다. 예를 들어 X가 확률변수인지 확인할 때 모든 Borel 집합 B에 대해 $X^{-1}(B) \in \mathcal{F}$를 확인할 필요 없이, 반직선 $(-\infty, b]$들에 대해서만 확인해도 충분하다(다음 절).

7.4 컴팩트 집합(Heine–Borel)

Definition 7.8 (컴팩트)**.** $\mathbb{R}^n$에서 집합 K가 닫혀 있고 유계이면 K를 컴팩트(compact)라 한다.

Theorem 7.9 (Heine–Borel)**.** *$\mathbb{R}^n$에서 집합이 컴팩트일 필요충분조건은 닫혀 있고 유계인 것이다.*

Remark 7.10. 컴팩트성은 확률론에서 *tightness*와 직결되며, 약수렴(weak convergence)의 핵심 정리(Prokhorov 정리)에서 중요한 역할을 한다.

8 Random variable

8.1 정의

Definition 8.1 (확률변수)**.** 확률공간 $(\Omega, \mathcal{F}, \mathbb{P})$가 주어졌다고 하자. 함수 $X : \Omega \to \mathbb{R}$가 다음을 만족하면 X를 *확률변수*(random variable)라 한다:

$$\forall B \in \mathcal{B}(\mathbb{R}), \qquad X^{-1}(B) \in \mathcal{F}.$$

즉, X는 $(\Omega, \mathcal{F})$에서 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$로 가는 가측함수이다.

Remark 8.2. 역상 $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$는 사건(event)이므로 확률 $\mathbb{P}(X \in B)$를 정의할 수 있다.

8.2 가측성 판정: 반직선만 확인해도 충분

Theorem 8.3 (반직선으로 판정)**.** *함수 $X : \Omega \to \mathbb{R}$에 대하여 다음은 동치이다.*

(1) *X는 확률변수이다.*

(2) *모든 $b \in \mathbb{R}$에 대해 $\{\omega : X(\omega) \le b\} = X^{-1}((-\infty, b]) \in \mathcal{F}$이다.*

구조화된 증명 **(1)** (1)⇒(2)는 정의에서 자명하다.

(2) (2)⇒(1)을 보이자.

(2.1) $\mathcal{H}_c := \{(-\infty, b] : b \in \mathbb{R}\}$라 두면 가정은 $X^{-1}(H) \in \mathcal{F}$가 모든 $H \in \mathcal{H}_c$에 대해 성립한다는 뜻이다.

(2.2) Theorem 7.6에 의해 $\sigma(\mathcal{H}_c) = \mathcal{B}(\mathbb{R})$이다.

(2.3) 역상 연산은 가산 합집합/여집합을 보존한다(Functions 절의 Proposition 참고). 따라서 $\mathcal{D} := \{B \subseteq \mathbb{R} : X^{-1}(B) \in \mathcal{F}\}$는 σ-algebra이고 $\mathcal{H}_c \subseteq \mathcal{D}$이다.

(2.4) 그러므로 $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{H}_c) \subseteq \mathcal{D}$. 즉 모든 $B \in \mathcal{B}(\mathbb{R})$에 대해 $X^{-1}(B) \in \mathcal{F}$, 곧 X는 확률변수이다.

□

8.3 기본 예시: 지시확률변수와 단순확률변수

Definition 8.4 (지시확률변수)**.** 사건 $A \in \mathcal{F}$에 대해

$$\mathbf{1}_A(\omega) := \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A \end{cases}$$

를 *지시확률변수*(indicator random variable)라 한다.

Definition 8.5 (단순확률변수)**.** 서로소 사건 $A_1, \dots, A_k \in \mathcal{F}$와 실수 $a_1, \dots, a_k$가 주어졌을 때

$$X(\omega) = \sum_{i=1}^{k} a_i \mathbf{1}_{A_i}(\omega)$$

꼴의 확률변수를 *단순확률변수*(simple r.v.)라 한다. (필요하면 $\{A_i\}$가 Ω의 분할(partition)이 되도록 $A_0 := \Omega \setminus \cup_{i=1}^{k} A_i$를 추가할 수 있다.)

8.4 가측함수의 합성과 연산 닫힘성

Theorem 8.6 (가측함수 합성)**.** *X가 확률변수이고 $f : \mathbb{R} \to \mathbb{R}$가 Borel-가측이면 $f \circ X$도 확률변수이다. 특히 연속함수는 Borel-가측이므로 연속함수 f에 대해 $f(X)$는 확률변수이다.*

구조화된 증명 **(1)** 임의의 Borel 집합 $B \in \mathcal{B}(\mathbb{R})$에 대해

$$(f \circ X)^{-1}(B) = X^{-1}(f^{-1}(B)).$$

(2) f가 Borel-가측이므로 $f^{-1}(B) \in \mathcal{B}(\mathbb{R})$이고, X가 확률변수이므로 $X^{-1}(f^{-1}(B)) \in \mathcal{F}$이다.

□

Theorem 8.7 (사칙연산 및 최대/최소)**.** *X, Y가 확률변수이면 다음도 모두 확률변수이다: $aX + bY$ $(a, b \in \mathbb{R})$, XY, $|X|$, $\max(X, Y)$, $\min(X, Y)$.*

설명/아이디어

- $(X, Y) : \Omega \to \mathbb{R}^2$가 Borel-가측임(즉, 랜덤벡터임)을 먼저 확인한다.
- $\mathbb{R}^2$에서 $\phi(x, y) = ax + by$, $\psi(x, y) = xy$, $\eta(x, y) = \max(x, y)$ 등은 연속함수이므로 Borel-가측이다.
- 따라서 합성 $\phi(X, Y)$, $\psi(X, Y)$, $\eta(X, Y)$는 모두 확률변수이다.

□

8.5 확률변수가 생성하는 σ-algebra

Definition 8.8 ($\sigma(X)$)**.** 확률변수 X가 생성하는 σ-algebra를

$$\sigma(X) := \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$$

로 정의한다.

Remark 8.9. $\sigma(X)$는 X를 확률변수로 만드는 가장 작은 σ-algebra이며, 직관적으로는 "X를 관측함으로써 얻는 모든 정보"를 담고 있다. 또한 Theorem 8.3에 의해

$$\sigma(X) = \sigma(\{X \le b\} : b \in \mathbb{R}).$$

9 Measure

이번 절에서는 $\mathbb{R}$ 위의 대표적인 측도 두 가지를 정리한다.

- 르베그 측도(Lebesgue measure) λ,
- 분포함수로부터 만들어지는 르베그–스틸체스 측도(Lebesgue–Stieltjes measure).

확률론에서 확률변수의 분포는 결국 $\mathbb{R}$ 위의 확률측도이며, 그 대표적인 예가 "밀도함수를 갖는 경우"(르베그 측도에 대한 절대연속)이다.

9.1 측도의 정의

Definition 9.1 (측도)**.** 가측공간 $(X, \mathcal{A})$에서 함수 $\mu : \mathcal{A} \to [0, \infty]$가

(1) $\mu(\emptyset) = 0$,

(2) 서로소인 $(A_n) \subseteq \mathcal{A}$에 대해 $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$

를 만족하면 μ를 (양의) 측도(measure)라 한다.

Remark 9.2. 확률측도는 $\mu(X) = 1$을 만족하는 특별한 측도이다.

9.2 Borel 측도와 르베그 측도(완비화)

Definition 9.3 (Borel 르베그 측도)**.** $\mathbb{R}$의 Borel σ-algebra $\mathcal{B}(\mathbb{R})$ 위에서

$$\lambda_B((a, b]) = b - a \qquad (a < b)$$

를 만족하는 측도 λ_B가 유일하게 존재한다. 이를 *(Borel) 르베그 측도*라 한다.

Remark 9.4. λ_B는 $\mathcal{B}(\mathbb{R})$ 위에서는 *완비(complete)*가 아니다. 즉, $\lambda_B(A) = 0$인 Borel 집합 A의 부분집합이 Borel 집합이 아닐 수 있다. (예: 칸토어 집합의 부분집합 중에는 비가측/비Borel인 것들이 존재할 수 있다.)

Definition 9.5 (르베그 σ-algebra와 르베그 측도)**.** λ_B를 완비화(completion)하여 얻는 σ-algebra를 르베그 σ-algebra $\mathcal{L}(\mathbb{R})$라 하고, 그 위로 확장된 완비 측도를 르베그 측도 λ라 한다. 즉,

$$\mathcal{L}(\mathbb{R}) := \sigma\big(\mathcal{B}(\mathbb{R}) \cup \mathcal{N}\big), \qquad \mathcal{N} := \{N \subseteq B : B \in \mathcal{B}(\mathbb{R}),\ \lambda_B(B) = 0\},$$

이며 λ는 $\mathcal{L}(\mathbb{R})$ 위에서 λ_B의 자연스러운 확장이다.

Theorem 9.6 (르베그 측도의 기본 성질)**.** *르베그 측도 λ는 다음을 만족한다.*

(1) 완비성*: $A \in \mathcal{L}(\mathbb{R})$, $\lambda(A) = 0$, $N \subseteq A \Rightarrow N \in \mathcal{L}(\mathbb{R})$.*

(2) σ-유한성*: $\mathbb{R} = \bigcup_{n \in \mathbb{N}} [-n, n]$이고 $\lambda([-n, n]) < \infty$.*

(3) 국소 유한성*: 모든 유계집합 K에 대해 $\lambda(K) < \infty$.*

(4) 평행이동 불변성*: $\lambda(A + x) = \lambda(A)$ $\quad (A + x := \{a + x : a \in A\})$.*

Definition 9.7 (almost everywhere)**.** 어떤 성질 $P(x)$가

$$\lambda(\{x \in \mathbb{R} : P(x) \text{ 가 성립하지 않음}\}) = 0$$

이면 "P가 거의 어디서나(almost everywhere, a.e.) 성립한다"고 한다.

Example 9.8 (유리수는 르베그 영집합)**.** $\mathbb{Q}$는 가산 집합이므로 열거 $\mathbb{Q} = \{q_1, q_2, \dots\}$가 가능하다. 각 n에 대해 길이가 $\varepsilon 2^{-n}$인 구간 $I_n := (q_n - \varepsilon 2^{-(n+1)}, q_n + \varepsilon 2^{-(n+1)})$를 잡으면 $\mathbb{Q} \subseteq \bigcup_n I_n$이고

$$\lambda(\mathbb{Q}) \leq \lambda\Big(\bigcup_n I_n\Big) \leq \sum_n \lambda(I_n) = \sum_n \varepsilon 2^{-n} = \varepsilon.$$

임의의 $\varepsilon > 0$에 대해 $\lambda(\mathbb{Q}) \leq \varepsilon$이므로 $\lambda(\mathbb{Q}) = 0$이다.

9.3 분포함수와 르베그–스틸체스 측도

Definition 9.9 (분포함수(CDF))**.** 함수 $F : \mathbb{R} \to \mathbb{R}$가 다음을 만족하면 (일반화된) *분포함수*라 한다.

(1) F는 단조증가(non-decreasing) 함수이다.

(2) F는 오른쪽 연속(right-continuous)이다: $\lim_{x \downarrow x_0} F(x) = F(x_0)$.

확률변수의 CDF는 추가로 $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to +\infty} F(x) = 1$을 만족한다.

Theorem 9.10 (르베그–스틸체스 측도)**.** *분포함수 F가 주어지면, Borel 집합 위에 유일한 측도 μ_F가 존재하여 모든 $a < b$에 대해*

$$\mu_F((a, b]) = F(b) - F(a)$$

를 만족한다. 이를 르베그–스틸체스 측도*라 한다.*

9.4 push-forward(유도) 측도와 분포

확률공간 $(\Omega, \mathcal{F}, \mathbb{P})$와 확률변수 $X : \Omega \to \mathbb{R}$가 주어졌다고 하자.

$$\mu_X(B) := \mathbb{P}(X \in B) = \mathbb{P}(X^{-1}(B)), \qquad B \in \mathcal{B}(\mathbb{R})$$

로 정의되는 측도 μ_X를 X의 *분포*(distribution) 또는 *유도 측도*(push-forward measure)라 한다. 즉, $\mu_X = \mathbb{P} \circ X^{-1}$이다.

$F_X(b) := \mathbb{P}(X \leq b)$는 X의 CDF이며, 이 F_X로부터 만들어진 르베그–스틸체스 측도 μ_{F_X}는 μ_X와 일치한다.

10 Sequence of functions

함수열 $(f_n)_{n \in \mathbb{N}}$은 모든 자연수 n에 대해, f_n이 공통된 정의역과 공역을 가진 함수이다.

Def. 점별 수렴 함수열 $(f_n)_{n \in \mathbb{N}}$이 모든 점 x에 대해, 다음이 성립하면 '점별 수렴'이라고 한다.

$$\lim_{n \to \infty} f_n(x) = f(x)$$

즉, 점별 수렴은 각 점마다 자연수 N의 값이 다를 수 있다.

Def. 균등 수렴 함수열 $(f_n)_{n \in \mathbb{N}}$이 임의의 $\epsilon > 0$에 대해, 어떤 자연수 N이 존재하여, 모든 점 x에서 $|f_n(x) - f(x)| < \epsilon$이 성립한다면, (f_n)은 f로 균등 수렴한다고 한다.

Theorem 10.1. *함수열 f_n이 f로 균등 수렴한다고 하자. 그러면, f_n은 f로 점별 수렴한다. 그 역은 성립하지 않는다.*

Def. Almost everywhere convergence (거의 어디서나 수렴**)** : 함수열 (f_n)이 함수 f로 거의 어디서나 (a.e.) 수렴한다 함은, 다음을 만족하는 null set N이 존재한다는 것이다:

$$\mu(N) = 0 \quad \text{and} \quad \forall x \notin N, \ \lim_{n \to \infty} f_n(x) = f(x).$$

확률론에서 이것은 almost sure convergence에 해당한다.

11 Lebesgue integral

확률론에서 기댓값은 단순히 "평균"이 아니라 *측도에 대한 적분*이다:

$$\mathbb{E}[X] = \int_\Omega X \, d\mathbb{P}.$$

따라서 확률론의 엄밀한 기초는 르베그 적분(=측도론적 적분)에 있다. 르베그 적분은

(1) 적분 가능한 함수의 범위를 크게 확장하고,

(2) 극한과 적분의 교환을 (MCT/DCT 등의) 자연스러운 조건 아래에서 가능하게 하며,

(3) 다중 적분의 순서 교환(Fubini–Tonelli)을 체계적으로 다룬다.

11.1 정의의 단계

르베그 적분(기댓값)을 "단순함수 → 일반함수"로 확장하는 방식으로 정의한다.

Step 1. 지시함수의 적분:

$$\int_\Omega \mathbf{1}_A \, d\mathbb{P} = \mathbb{P}(A) \qquad (A \in \mathcal{F}).$$

Step 2. (음이 아닌) 단순확률변수 $X(\omega) = \sum_{i=1}^k a_i \mathbf{1}_{A_i}(\omega)$ $(a_i \geq 0)$의 적분:

$$\int_\Omega X \, d\mathbb{P} = \sum_{i=1}^k a_i \mathbb{P}(A_i).$$

Step 3. (단순함수 근사 정리)

Theorem 11.1 (단순함수 근사)**.** *$X : \Omega \to [0, \infty]$가 가측이면, $0 \leq X_1 \leq X_2 \leq \cdots$ 이고 $X_n(\omega) \uparrow X(\omega)$ (모든 ω)인 음이 아닌 단순확률변수열 (X_n)이 존재한다.*

구조화된 증명 **(1)** 각 $n \in \mathbb{N}$에 대해 다음과 같이 X_n을 정의한다:

$$X_n(\omega) := \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbf{1}_{\{\frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n}\}} \; + \; n \mathbf{1}_{\{X(\omega) \geq n\}}.$$

(2) 각 항은 가측집합의 지시함수이므로 X_n은 단순확률변수이며 $0 \leq X_n \leq X$이다.

(3) n이 증가하면 격자폭 2^{-n}이 작아지고 절단수준 n이 커지므로 $X_n(\omega) \leq X_{n+1}(\omega)$가 모든 ω에서 성립한다.

(4) 임의의 ω를 고정하자.

(4.1) 만약 $X(\omega) < \infty$이면, 큰 n에서는 $X(\omega) < n$이므로 마지막 항이 사라지고, 구간 분할의 격자폭이 2^{-n}으로 작아져 $X_n(\omega) \uparrow X(\omega)$가 된다.

(4.2) 만약 $X(\omega) = \infty$이면, 모든 n에 대해 마지막 항이 n이므로 $X_n(\omega) \uparrow \infty = X(\omega)$.

(5) 따라서 $X_n \uparrow X$.

□

Step 4. 음이 아닌 가측함수의 적분:

$$\int_\Omega X \, d\mathbb{P} := \sup \left\{ \int_\Omega \phi \, d\mathbb{P} : \; 0 \leq \phi \leq X, \; \phi \text{ 단순함수} \right\}.$$

Step 5. 일반 가측함수의 적분: $X = X^+ - X^-$ (양/음의 부분)으로 두고,

$$\int_\Omega X \, d\mathbb{P} := \int_\Omega X^+ \, d\mathbb{P} - \int_\Omega X^- \, d\mathbb{P}$$

로 정의한다. 이 값이 유한하려면 $\int |X| \, d\mathbb{P} < \infty$가 필요하다.

Definition 11.2 (적분가능(=기댓값 존재))**.** $\mathbb{E}[|X|] < \infty$이면 X를 *적분가능(integrable)*하다고 한다.

11.2 핵심 수렴 정리들: Fatou, MCT, DCT

Theorem 11.3 (Fatou's Lemma)**.** *음이 아닌 가측함수열* (X_n)*에 대해*

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \le \liminf_{n\to\infty} \mathbb{E}[X_n].$$

구조화된 증명 **(1)** $Y_n := \inf_{k\ge n} X_k$라 두면 Y_n은 가측이고 $Y_n \uparrow \liminf_{n\to\infty} X_n$이다.

(2) 임의의 음이 아닌 단순함수 $\phi \le \liminf X_n$를 고정하자. 각 n에 대해

$$B_n := \{\omega : \ \phi(\omega) \le Y_n(\omega)\}$$

라 두면 $B_n \uparrow \Omega$이고, B_n 위에서는 $\phi \le Y_n \le X_k$ (모든 $k \ge n$)이다.

(3) 단순함수의 적분(Step 2)과 측도의 연속성으로

$$\int \phi \, d\mathbb{P} = \lim_{n\to\infty} \int \phi \mathbf{1}_{B_n} \, d\mathbb{P}.$$

(4) 각 n에 대해 $\phi \mathbf{1}_{B_n} \le X_k$ $(k \ge n)$이므로 단조성으로

$$\int \phi \mathbf{1}_{B_n} \, d\mathbb{P} \le \inf_{k\ge n} \int X_k \, d\mathbb{P}.$$

(5) $n \to \infty$를 취하면

$$\int \phi \, d\mathbb{P} \le \liminf_{k\to\infty} \int X_k \, d\mathbb{P}.$$

(6) 마지막으로 $\phi \le \liminf X_n$인 모든 단순함수 ϕ에 대해 성립하므로, Step 4(상한 정의)에 의해 $\int \liminf X_n \, d\mathbb{P} \le \liminf \int X_n \, d\mathbb{P}$.

□

Theorem 11.4 (Monotone Convergence Theorem (MCT))**.** $0 \le X_1 \le X_2 \le \cdots$ *이고* $X_n \uparrow X$ *이면*

$$\lim_{n\to\infty} \int X_n \, d\mathbb{P} = \int X \, d\mathbb{P}.$$

구조화된 증명 **(1)** 단조성으로 $\int X_n \, d\mathbb{P} \le \int X \, d\mathbb{P}$이므로 $\limsup_n \int X_n \, d\mathbb{P} \le \int X \, d\mathbb{P}$.

(2) 한편 Fatou 보조정리(Theorem 11.3)를 X_n에 적용하면

$$\int X \, d\mathbb{P} = \int \liminf_n X_n \, d\mathbb{P} \le \liminf_n \int X_n \, d\mathbb{P}.$$

(3) 두 부등식을 합치면 $\lim_n \int X_n \, d\mathbb{P} = \int X \, d\mathbb{P}$.

□

Theorem 11.5 (Dominated Convergence Theorem (DCT))**.** $X_n \to X$ *a.s.이고 어떤 적분가능한* Y*가 존재하여* $|X_n| \le Y$ *a.s.이면*

$$\lim_{n\to\infty} \int X_n \, d\mathbb{P} = \int X \, d\mathbb{P}.$$

특히 $\mathbb{E}[|X_n - X|] \to 0$.

구조화된 증명 **(1)** $|X_n| \le Y$이고 $X_n \to X$이므로 $|X| \le Y$ a.s.이고 X도 적분가능하다.

(2) 음이 아닌 함수열 $Y + X_n$에 Fatou를 적용하면

$$\int (Y + X) \, d\mathbb{P} \le \liminf_n \int (Y + X_n) \, d\mathbb{P} = \int Y \, d\mathbb{P} + \liminf_n \int X_n \, d\mathbb{P},$$

따라서 $\int X \, d\mathbb{P} \le \liminf_n \int X_n \, d\mathbb{P}$.

(3) 같은 방식으로 $Y - X_n$에 Fatou를 적용하면 $\int X\,d\mathbb{P} \geq \limsup_n \int X_n\,d\mathbb{P}$를 얻는다.

(4) 결론적으로 $\int X_n\,d\mathbb{P} \to \int X\,d\mathbb{P}$.

(5) 또한 $|X_n - X| \leq 2Y$이고 $|X_n - X| \to 0$ a.s.이므로 위 결과를 $|X_n - X|$에 적용해 $\mathbb{E}[|X_n - X|] \to 0$을 얻는다.

□

11.3 Fubini–Tonelli 정리

확률론에서 결합분포/주변분포를 다룰 때 다중 적분의 순서 교환이 핵심이며, 이는 곱측도와 Fubini–Tonelli로 정당화된다.

Theorem 11.6 (곱측도의 존재(진술)). *$(\Omega_1, \mathcal{F}_1, \mu_1)$, $(\Omega_2, \mathcal{F}_2, \mu_2)$가 σ-finite 측도공간이면, 곱 σ-algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ 위에 유일한 측도 $\mu_1 \times \mu_2$가 존재하여*

$$(\mu_1 \times \mu_2)(A \times B) = \mu_1(A)\mu_2(B) \qquad (A \in \mathcal{F}_1,\ B \in \mathcal{F}_2)$$

를 만족한다.

Theorem 11.7 (Tonelli). *Theorem 11.6의 설정에서 $f : \Omega_1 \times \Omega_2 \to [0, \infty]$가 가측이면*

$$\int_{\Omega_1 \times \Omega_2} f\,d(\mu_1 \times \mu_2) = \int_{\Omega_1} \left(\int_{\Omega_2} f(x, y)\,d\mu_2(y) \right) d\mu_1(x) = \int_{\Omega_2} \left(\int_{\Omega_1} f(x, y)\,d\mu_1(x) \right) d\mu_2(y).$$

구조화된 증명(개요). **(1)** f가 직사각형 지시함수 $\mathbf{1}_{A \times B}$이면 등식은 곱측도의 정의로부터 성립한다.

(2) f가 음이 아닌 단순함수이면 선형성으로 성립한다.

(3) 일반 $f \geq 0$에 대해 Theorem 11.1로 $f_n \uparrow f$인 단순함수 근사를 취하고, MCT(Theorem 11.4)로 극한을 취해 결론을 얻는다.

□

Theorem 11.8 (Fubini). *Theorem 11.6의 설정에서 f가 적분가능($\int |f|\,d(\mu_1 \times \mu_2) < \infty$)이면 Tonelli의 등식이 그대로 성립한다.*

구조화된 증명(개요). **(1)** $f = f^+ - f^-$로 분해하고 $f^\pm$에 Tonelli(Theorem 11.7)를 적용한다.

(2) 적분가능성 가정은 $\int f^+ < \infty$와 $\int f^- < \infty$를 보장하므로 두 값의 차가 잘 정의된다.

(3) 따라서 반복적분이 존재하고 값이 곱적분과 같다.

□

Corollary 11.9 (꼬리적분 공식(layer-cake)). *$X \geq 0$이면*

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > t)\,dt.$$

구조화된 증명 **(1)** $X(\omega) = \int_0^\infty \mathbf{1}_{\{X(\omega) > t\}}\,dt$가 성립한다. (고정된 ω에서 구간 길이 계산)

(2) Tonelli 정리(Theorem 11.7)로 적분 순서를 교환하면

$$\mathbb{E}[X] = \int_\Omega \left(\int_0^\infty \mathbf{1}_{\{X > t\}}\,dt \right) d\mathbb{P} = \int_0^\infty \left(\int_\Omega \mathbf{1}_{\{X > t\}}\,d\mathbb{P} \right) dt = \int_0^\infty \mathbb{P}(X > t)\,dt.$$

□

Remark 11.10 (LOTUS: 분포로 적분하기). 분포측도 $\mu_X = \mathbb{P} \circ X^{-1}$를 쓰면

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)\,\mu_X(dx)$$

이며, 이는 "변수 치환"의 측도론적 형태로서 확률론 계산의 표준 도구다.

12 Discrete or Continuous Random variable

확률변수 X의 분포는 CDF $F_X(b) = \mathbb{P}(X \le b)$로 완전히 결정된다. 하지만 "연속/이산" 이라는 용어는 문맥에 따라 서로 다른 의미로 쓰일 수 있으므로, 측도론적 관점에서 분포를 세 가지로 구분해 두는 것이 가장 안전하다.

12.1 원자(atom)와 이산 분포

Definition 12.1 (원자(atomic point))**.** 점 $x \in \mathbb{R}$가 X의 *원자*(atom)라 함은 $\mathbb{P}(X = x) > 0$인 것을 말한다.

Definition 12.2 (이산(discrete) 확률변수)**.** X가 *이산*이라 함은 X의 가능한 값들의 집합(지지집합)

$$S := \{x \in \mathbb{R} : \mathbb{P}(X = x) > 0\}$$

이 가산 집합인 경우를 말한다.

이산 확률변수에서는 *확률질량함수*(PMF)

$$p(x) := \mathbb{P}(X = x) \qquad (x \in S)$$

로 분포가 표현되며,

$$\mathbb{P}(X \in A) = \sum_{x \in A \cap S} p(x)$$

이다. 또한 $\sum_{x \in S} |x| p(x) < \infty$이면

$$\mathbb{E}[X] = \sum_{x \in S} x\, p(x).$$

12.2 연속성과 절대연속성

Definition 12.3 (연속 분포(continuous distribution))**.** F_X가 연속함수이면(즉 모든 점에서 점프가 없으면) X는 *연속 분포*를 가진다고 한다. 이 경우 모든 $a \in \mathbb{R}$에 대해 $\mathbb{P}(X = a) = 0$이다.

Remark 12.4. 연속 분포라고 해서 항상 밀도함수(PDF)가 존재하는 것은 아니다. 예를 들어 "칸토어 분포"는 CDF가 연속이지만 르베그 측도에 대해 특이(singular)하여 밀도함수가 없다. 통계학에서는 보통 "연속 확률변수"를 "밀도함수를 가진 경우(절대연속)"로 사용하는 경우가 많으니, 문맥에 따라 구분해 읽어야 한다.

Definition 12.5 (절대연속(absolute continuity)과 밀도(PDF))**.** X의 분포측도 μ_X가 르베그 측도 λ에 대해 절대연속($\mu_X \ll \lambda$)이면, Radon–Nikodym 정리에 의해 (a.e. 유일한) 함수 $f \ge 0$가 존재하여

$$\mu_X(A) = \int_A f(x)\, d\lambda(x) \qquad (A \in \mathcal{B}(\mathbb{R})).$$

이 $f = \frac{d\mu_X}{d\lambda}$를 X의 *확률밀도함수*(PDF)라 한다.

이때

$$F_X(x) = \int_{(-\infty,x]} f(t)\, dt = \int_{-\infty}^{x} f(t)\, dt, \qquad \mathbb{E}[X] = \int_{\mathbb{R}} x f(x)\, dx$$

(적분 가능하다고 가정).

12.3 혼합(mixed) 분포

일반적인 분포는 "이산 + 절대연속 + 특이연속" 성분을 동시에 가질 수 있다. 즉, 하나의 CDF는 점프(원자) 부분과 연속 부분이 함께 존재할 수 있다. 확률론/통계학에서 "이산/연속"을 말할 때는 보통 이러한 점프 유무 혹은 밀도 존재 여부를 염두에 둔다.

12.4 Radon–Nikodym 정리(밀도의 존재와 유일성)

Radon–Nikodym 정리는 "한 측도가 다른 측도에 대해 절대연속이면, 그 비율이 점wise 밀도(density)로 표현된다"는 사실을 엄밀히 정식화한다. 확률론에서는 PDF, 조건부기댓값, 베이즈 정리의 측도론적 형태 등이 모두 이 정리에 기반한다.

Theorem 12.6 (Radon–Nikodym)**.** *측도공간 $(S, \mathcal{A})$에서 μ, ν가 σ-finite measure라 하자. 만약 $\nu \ll \mu$이면 어떤 $\mathcal{A}$-가측 함수 $f : S \to [0, \infty]$가 존재하여*

$$\nu(A) = \int_A f \, d\mu \qquad (\forall A \in \mathcal{A})$$

가 성립한다. 이러한 f는 μ-a.e. 유일하며 $f = \frac{d\nu}{d\mu}$로 쓴다.

Remark 12.7 (확률론에서의 해석). **(1)** $\mu = \lambda$ (르베그 측도), $\nu = \mu_X$ (분포측도)인 경우 $\frac{d\mu_X}{d\lambda}$가 PDF 이다.

(2) $(\Omega, \mathcal{F}, \mathbb{P})$에서 부분 σ-algebra $\mathcal{G}$와 적분가능한 X에 대해, $\nu(A) := \int_A X \, d\mathbb{P}$ $(A \in \mathcal{G})$로 두면 $\nu \ll \mathbb{P}|_{\mathcal{G}}$이고 Radon–Nikodym 정리로부터 $\mathcal{G}$-가측인 $\mathbb{E}[X \mid \mathcal{G}]$가 존재한다. (이는 조건부기댓값의 존재/유일성의 핵심이다.)

증명(아이디어). 정리의 완전한 증명은 측도론의 핵심 난이도를 갖는다(하ーン 분해, signed measure 등). 여기서는 확률론에서 사용될 때의 흐름이 보이도록 구조만 요약한다.

(1) σ-finite 가정 하에 공간을 μ-유한한 조각들로 분해하여 "유한 측도" 경우로 환원한다.

(2) 유한 측도인 경우, ν와 μ의 차이를 signed measure로 보고 "최대한 ν를 잘 근사하는" 밀도 후보를 최적화(상한 달성)로 구성한다.

(3) 그 후보 f에 대해 $\nu(A) - \int_A f \, d\mu$가 모든 A에서 0이 됨을 보여 $\nu(A) = \int_A f \, d\mu$를 얻는다.

(4) 유일성은 만약 f, g가 둘 다 위 등식을 만족하면 $\int_A (f - g) \, d\mu = 0$가 모든 A에 대해 성립하고, 특히 $A = \{f > g\}, \{g > f\}$를 대입하여 $\mu(\{f \neq g\}) = 0$을 얻는다.

□

13 Moment

13.1 모멘트, 기댓값, 분산

확률변수의 *모멘트(moment)*는 분포의 "크기/형태"를 수치로 요약해 주는 도구다. 특히 1차 모멘트는 평균, 2차 중심모멘트는 분산이며, 고차 모멘트는 꼬리의 두께나 비대칭성을 정량화한다.

Definition 13.1 (모멘트와 중심 모멘트)**.** 적분가능한 확률변수 X와 정수 $n \geq 1$에 대해

$$m_n := \mathbb{E}[X^n] \quad (\text{존재하면 } n\text{차 모멘트}), \qquad \mu_n := \mathbb{E}[(X - \mathbb{E}[X])^n] \quad (\text{존재하면 } n\text{차 중심 모멘트})$$

라 한다.

Definition 13.2 (분산과 공분산)**.** $X \in L^2$이면

$$\operatorname{Var}(X) := \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

를 *분산*이라 한다. 또한 $X, Y \in L^2$이면

$$\operatorname{Cov}(X, Y) := \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

를 *공분산*이라 한다.

Remark 13.3 (해석학적 의미). L^2에서 분산은 "평균으로부터의 거리"의 제곱에 해당한다:

$$\sqrt{\operatorname{Var}(X)} = \|X - \mathbb{E}[X]\|_2.$$

따라서 분산을 최소화하는 문제는 L^2-기하(정사영)로 해석되며, 이는 뒤의 *조건부 기댓값* $=$ L^2 *정사영*으로 이어진다.

13.2 평균은 평균제곱오차를 최소화한다

Theorem 13.4 (평균의 최적성)**.** *$X \in L^2$이면 임의의 $a \in \mathbb{R}$에 대해*

$$\mathbb{E}[(X-a)^2] \geq \mathbb{E}[(X-\mathbb{E}[X])^2] = \mathrm{Var}(X)$$

이며 등호는 $a = \mathbb{E}[X]$일 때(그리고 그때만) 성립한다.

구조화된 증명 목표**.** $\mathbb{E}[(X-a)^2] = \mathrm{Var}(X) + (\mathbb{E}[X]-a)^2$를 보이자.

(1) $X - a = (X - \mathbb{E}[X]) + (\mathbb{E}[X] - a)$로 분해한다.

(2) 제곱하여 전개하면

$$(X-a)^2 = (X-\mathbb{E}[X])^2 + 2(\mathbb{E}[X]-a)(X-\mathbb{E}[X]) + (\mathbb{E}[X]-a)^2.$$

(3) 양변에 기댓값을 취한다.

(3.1) $\mathbb{E}[X-\mathbb{E}[X]] = 0$이므로 교차항의 기댓값은 $2(\mathbb{E}[X]-a)\mathbb{E}[X-\mathbb{E}[X]] = 0$이다.

(3.2) 따라서 $\mathbb{E}[(X-a)^2] = \mathbb{E}[(X-\mathbb{E}[X])^2] + (\mathbb{E}[X]-a)^2 = \mathrm{Var}(X) + (\mathbb{E}[X]-a)^2$.

(4) 마지막 항 $(\mathbb{E}[X]-a)^2 \geq 0$이므로 부등식이 성립하고, 등호는 $a = \mathbb{E}[X]$일 때만 가능하다.

□

13.3 생성함수: MGF와 cumulant

Definition 13.5 (적률생성함수(MGF))**.** 확률변수 X에 대해

$$M_X(\lambda) := \mathbb{E}[e^{\lambda X}]$$

가 어떤 λ의 근방에서 유한하면 M_X를 *적률생성함수(moment generating function)*라 한다. M_X가 존재하는 구간에서는

$$M_X^{(n)}(0) = \mathbb{E}[X^n]$$

이 성립한다.

Remark 13.6 (왜 MGF가 유용한가). MGF는 독립합에 대해 곱으로 분해되며($M_{X+Y} = M_X M_Y$), 테일 확률을 지수적으로 제어하는 Chernoff bound의 핵심 입력이다. 반면 모든 분포에서 존재하는 것은 아니다 (예: 코시 분포).

Theorem 13.7 (MGF의 유일성)**.** *두 확률변수 X, Y의 MGF가 0의 열린 근방에서 존재하고 그 근방에서 $M_X \equiv M_Y$이면 X와 Y는 같은 분포를 가진다.*

구조화된 증명(개요). **(1)** MGF가 0의 근방에서 유한하면 그 근방에서 해석함수(거듭제곱급수)로 전개된다.

(2) $M_X \equiv M_Y$이면 모든 도함수 값이 일치하므로 $\mathbb{E}[X^n] = \mathbb{E}[Y^n]$가 모든 n에 대해 성립한다.

(3) 한편 MGF는 라플라스 변환과 동일하며, 라플라스 변환의 유일성 정리에 의해 분포가 결정된다. (보다 표준적으로는 $M_X(\lambda) = \varphi_X(-i\lambda)$로 특성함수의 해석적 연장과 연결하여 증명한다.)

□

13.4 기본 부등식과 꼬리 확률

이 절의 부등식들은 이후 *집중 부등식(concentration)*의 출발점이다.

Theorem 13.8 (Markov 부등식)**.** *음이 아닌 확률변수 $X \geq 0$와 $a > 0$에 대해*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

구조화된 증명 **(1)** 사건 $\{X \geq a\}$에서는 $X \geq a\mathbf{1}_{\{X\geq a\}}$가 성립한다.

(2) 양변에 기댓값을 취하면 $\mathbb{E}[X] \geq a\,\mathbb{E}[\mathbf{1}_{\{X\geq a\}}] = a\,\mathbb{P}(X \geq a)$.

(3) 정리하면 원하는 부등식을 얻는다.

□

Theorem 13.9 (Chebyshev 부등식)**.** *$X \in L^2$와 $a > 0$에 대해*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}.$$

구조화된 증명 **(1)** $Y := (X - \mathbb{E}[X])^2 \geq 0$에 Markov 부등식(Theorem 13.8)을 적용한다.

(2) $\mathbb{P}(|X - \mathbb{E}[X]| \geq a) = \mathbb{P}((X - \mathbb{E}[X])^2 \geq a^2) \leq \mathbb{E}[(X - \mathbb{E}[X])^2]/a^2 = \mathrm{Var}(X)/a^2$.

□

Theorem 13.10 (Jensen 부등식)**.** *볼록함수 $\varphi : \mathbb{R} \to \mathbb{R}$와 적분가능한 확률변수 X에 대해*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$$

가 성립한다. (오목함수인 경우 부등호는 반대.)

구조화된 증명(접선 지지선). **(1)** 볼록함수의 성질에 의해 각 x_0에서 φ를 지지하는 접선(또는 부분기울기) $l_{x_0}(x) = \alpha x + \beta$가 존재하여 $l_{x_0}(x) \leq \varphi(x)$가 모든 x에 대해 성립하고, $l_{x_0}(x_0) = \varphi(x_0)$이다.

(2) $x_0 = \mathbb{E}[X]$로 두고 위 부등식에 기댓값을 취하면

$$\mathbb{E}[l_{\mathbb{E}[X]}(X)] \leq \mathbb{E}[\varphi(X)].$$

(3) 선형성으로 $\mathbb{E}[l_{\mathbb{E}[X]}(X)] = \alpha\mathbb{E}[X] + \beta = l_{\mathbb{E}[X]}(\mathbb{E}[X]) = \varphi(\mathbb{E}[X])$.

(4) 따라서 $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$.

□

Theorem 13.11 (Chernoff bound)**.** *임의의 확률변수 X와 임의의 $\lambda > 0$에 대해*

$$\mathbb{P}(X \geq t) \leq e^{-\lambda t}\mathbb{E}[e^{\lambda X}].$$

구조화된 증명 **(1)** 지수함수의 단조증가성을 이용하면 $\{X \geq t\} = \{e^{\lambda X} \geq e^{\lambda t}\}$.

(2) Markov 부등식을 $e^{\lambda X}$에 적용하면 $\mathbb{P}(e^{\lambda X} \geq e^{\lambda t}) \leq \mathbb{E}[e^{\lambda X}]/e^{\lambda t}$.

(3) 정리하면 원하는 부등식을 얻는다.

□

Remark 13.12 (통계/학습이론에서의 적용). Chebyshev는 분산만으로 느슨하지만 매우 일반적인 꼬리 상한을 주며, Chernoff는 MGF(또는 sub-Gaussian 성질)를 결합해 *지수 꼬리*를 얻는다. 후자는 유한표본 오차를 다루는 현대 통계학/기계학습 이론에서 핵심이다.

14 Characteristic function (특성함수)

MGF는 항상 존재하지 않으므로(예: 코시 분포), 모든 분포에서 항상 정의되는 변환이 필요하다. *특성함수 (characteristic function)*는 그 역할을 하며, 분포 수렴을 다루는 강력한 도구이기도 하다.

Definition 14.1 (특성함수)**.** 확률변수 X의 특성함수는

$$\varphi_X(t) := \mathbb{E}[e^{itX}], \qquad t \in \mathbb{R}$$

로 정의된다. 여기서 $i = \sqrt{-1}$이다.

Theorem 14.2 (기본 성질)**.** *임의의 확률변수 X에 대해 다음이 성립한다.*

(1) $\varphi_X(0) = 1$*이고* $|\varphi_X(t)| \le 1$*이다.*

(2) φ_X*는 균등연속(uniformly continuous)이다.*

(3) X, Y*가 독립이면* $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$*.*

(4) $\mathbb{E}[|X|^n] < \infty$*이면* φ_X*는* n*번 미분가능하고* $\varphi_X^{(n)}(0) = i^n\mathbb{E}[X^n]$*.*

구조화된 증명(대표 항목). **(1)** (1) $|e^{itX}| = 1$이므로 $|\varphi_X(t)| = |\mathbb{E}[e^{itX}]| \le \mathbb{E}[|e^{itX}|] = 1$이다. 또한 $\varphi_X(0) = \mathbb{E}[1] = 1$.

(2) (2) 임의의 s, t에 대해

$$|\varphi_X(t) - \varphi_X(s)| \le \mathbb{E}[|e^{itX} - e^{isX}|].$$

적분가능한 지배함수 2 아래에서 $t \to s$이면 integrand가 점wise로 0으로 가므로(DCT) 연속성. 더 나아가 $|e^{iu} - 1| \le |u|$를 이용하면

$$|\varphi_X(t) - \varphi_X(s)| \le \mathbb{E}[|X||t - s|]$$

가 $X \in L^1$일 때 성립하고, 일반 X에 대해서도 절단(truncation)으로 균등연속성을 얻는다.

(3) (3) 독립이면

$$\varphi_{X+Y}(t) = \mathbb{E}[e^{it(X+Y)}] = \mathbb{E}[e^{itX}e^{itY}] = \mathbb{E}[e^{itX}]\,\mathbb{E}[e^{itY}]$$

이며 마지막 등호는 독립성에 의한 곱 적분 분해(뒤의 Independence 절의 Lemma 15.5)로부터 따른다. □

Theorem 14.3 (유일성 정리)**.** $\varphi_X \equiv \varphi_Y$ *이면* X*와* Y*는 같은 분포를 가진다.*

Remark 14.4. 위 유일성은 *특성함수가 분포를 완전히 결정*함을 뜻한다. 이는 CLT 증명의 표준 경로(특성함수 $\to$ 점wise 수렴 $\to$ 분포수렴)를 가능하게 한다.

Theorem 14.5 (Lévy의 연속성 정리)**.** *확률변수열* (X_n)*과 확률변수* X*에 대해 다음은 동치이다.*

(1) $X_n \xrightarrow{d} X$*.*

(2) *모든* $t \in \mathbb{R}$*에 대해* $\varphi_{X_n}(t) \to \varphi_X(t)$*.*

Remark 14.6 (CLT에서의 역할). Theorem 14.5는 "특성함수의 점wise 수렴 $\Rightarrow$ 분포 수렴"을 보장한다. 따라서 정규분포의 특성함수 $e^{-t^2/2}$를 목표로 삼아 CLT를 증명할 수 있다.

15 Independence

15.1 독립의 세 가지 층: 사건, 시그마 대수, 확률변수

Definition 15.1 (사건의 독립)**.** 사건 $A, B \in \mathcal{F}$가 독립이라 함은 $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$인 것이다. 사건열 (A_n)이 (서로) 독립이라 함은 임의의 유한 부분집합 $\{n_1, \dots, n_k\}$에 대해

$$\mathbb{P}\Big(\bigcap_{j=1}^{k} A_{n_j}\Big) = \prod_{j=1}^{k} \mathbb{P}(A_{n_j})$$

가 성립하는 것이다.

Definition 15.2 (σ-algebra의 독립)**.** 부분 σ-algebra $\mathcal{G}_1, \dots, \mathcal{G}_n \subseteq \mathcal{F}$가 독립이라 함은 임의의 $A_i \in \mathcal{G}_i$에 대해

$$\mathbb{P}\Big(\bigcap_{i=1}^{n} A_i\Big) = \prod_{i=1}^{n} \mathbb{P}(A_i)$$

가 성립하는 것이다.

Definition 15.3 (확률변수의 독립)**.** 확률변수 $X_1, \dots, X_n$이 독립이라 함은 생성 σ-algebra $\sigma(X_1), \dots, \sigma(X_n)$이 독립인 것이다. 동치로, 임의의 Borel 집합 B_i에 대해

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i)$$

가 성립한다.

15.2 생성계로 독립을 확인하기: π-system 트릭

Theorem 15.4 (생성계 위에서의 독립 확인)**.** *각 i에 대해 $\mathcal{P}_i \subseteq \mathcal{F}$가 π-system이고 $\mathcal{G}_i := \sigma(\mathcal{P}_i)$라 하자. 만약 임의의 $A_i \in \mathcal{P}_i$에 대해*

$$\mathbb{P}\Big(\bigcap_{i=1}^{n} A_i\Big) = \prod_{i=1}^{n} \mathbb{P}(A_i)$$

가 성립하면, $\mathcal{G}_1, \dots, \mathcal{G}_n$은 서로 독립이다.

구조화된 증명(두 개의 경우). 표기 단순화를 위해 $n = 2$만 보인다. (일반 n도 동일한 귀납/반복으로 된다.)

(1) $\mathcal{L} := \{B \in \mathcal{G}_2 : \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\ \forall A \in \mathcal{P}_1\}$라 두자.

(2) $\mathcal{L}$이 λ-system임을 보이자. (확률의 가산가법성과 여집합 성질로 쉽게 확인된다.)

(3) 가정에 의해 $\mathcal{P}_2 \subseteq \mathcal{L}$이므로 Dynkin의 π-λ 정리로 $\mathcal{G}_2 = \sigma(\mathcal{P}_2) \subseteq \mathcal{L}$.

(4) 즉, 모든 $B \in \mathcal{G}_2$와 모든 $A \in \mathcal{P}_1$에 대해 $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

(5) 이제 고정된 $B \in \mathcal{G}_2$에 대해 $\mathcal{L}_B := \{A \in \mathcal{G}_1 : \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\}$라 두면, 같은 논리로 $\mathcal{L}_B$는 λ-system이고 $\mathcal{P}_1 \subseteq \mathcal{L}_B$이므로 $\mathcal{G}_1 \subseteq \mathcal{L}_B$.

(6) 따라서 모든 $A \in \mathcal{G}_1, B \in \mathcal{G}_2$에 대해 곱셈 공식이 성립하므로 $\mathcal{G}_1, \mathcal{G}_2$는 독립이다.

□

15.3 독립이면 곱의 기댓값이 분해된다

Lemma 15.5 (독립 ⇒ 곱의 기댓값 분해)**.** *X, Y가 독립이고 $X, Y \in L^1$이면 $XY \in L^1$이며*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

구조화된 증명(단순함수 근사). **(1)** 먼저 $X, Y \geq 0$이고 단순확률변수인 경우를 보이자.

(1.1) $X = \sum_{i=1}^{k} a_i \mathbf{1}_{A_i}$, $Y = \sum_{j=1}^{\ell} b_j \mathbf{1}_{B_j}$라 하자.

(1.2) 독립성으로 $\mathbb{P}(A_i \cap B_j) = \mathbb{P}(A_i)\mathbb{P}(B_j)$.

(1.3) 따라서

$$\mathbb{E}[XY] = \sum_{i,j} a_i b_j \mathbb{P}(A_i \cap B_j) = \sum_{i,j} a_i b_j \mathbb{P}(A_i)\mathbb{P}(B_j) = \mathbb{E}[X]\mathbb{E}[Y].$$

(2) 일반 $X, Y \geq 0$에 대해, 단순확률변수 증가열 $(X_n), (Y_n)$로 점wise 증가 근사한다(Theorem 11.1). 독립성은 $\sigma(X), \sigma(Y)$의 독립이므로 X_n, Y_n도 독립이다.

(3) 단조수렴정리(MCT, Theorem 11.4)로

$$\mathbb{E}[XY] = \lim_{n\to\infty} \mathbb{E}[X_n Y_n] = \lim_{n\to\infty} \mathbb{E}[X_n]\mathbb{E}[Y_n] = \mathbb{E}[X]\mathbb{E}[Y].$$

(4) 일반 부호의 경우 $X = X^+ - X^-$, $Y = Y^+ - Y^-$로 분해한 뒤 선형성을 적용한다.

□

Corollary 15.6 (독립인 합의 분산). *$X, Y \in L^2$가 독립이면*

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

구조화된 증명 **(1)** 정의에 의해 $\mathrm{Var}(X+Y) = \mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2$.

(2) 전개하면 $\mathbb{E}[X^2] + \mathbb{E}[Y^2] + 2\mathbb{E}[XY] - (\mathbb{E}[X] + \mathbb{E}[Y])^2$.

(3) Lemma 15.5로 $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$이므로 교차항이 상쇄되어 $\mathrm{Var}(X) + \mathrm{Var}(Y)$가 된다.

□

15.4 Kolmogorov의 0–1 법칙

Definition 15.7 (Tail σ-algebra). 확률변수열 $(X_n)_{n\geq 1}$에 대해

$$\mathcal{T} := \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \dots)$$

를 *꼬리 σ-algebra*라 한다. $\mathcal{T}$의 사건은 유한 개의 좌표를 바꾸어도 영향을 받지 않는 "장기적 사건"이다.

Theorem 15.8 (Kolmogorov 0–1 law). *(X_n)이 서로 독립이면 꼬리 σ-algebra $\mathcal{T}$의 모든 사건 A에 대해 $\mathbb{P}(A) \in \{0, 1\}$이다.*

구조화된 증명 **(1)** 각 n에 대해 $\mathcal{F}_n := \sigma(X_1, \dots, X_n)$라 두자. 독립성으로 $\mathcal{F}_n$과 $\sigma(X_{n+1}, X_{n+2}, \dots)$는 독립이다.

(2) $A \in \mathcal{T}$이면 특히 $A \in \sigma(X_{n+1}, X_{n+2}, \dots)$이므로 A는 $\mathcal{F}_n$과 독립이다. 즉, 모든 $B \in \mathcal{F}_n$에 대해 $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

(3) $\mathcal{F}_\infty := \sigma(\bigcup_{n\geq 1} \mathcal{F}_n) = \sigma(X_1, X_2, \dots)$라 두자. 또한 $\mathcal{P} := \bigcup_{n\geq 1} \mathcal{F}_n$는 증가열의 합집합이므로 유한 교집합에 대해 닫혀 있는 π-system이다.

(4) $\mathcal{L} := \{B \in \mathcal{F}_\infty : \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\}$라 두자. 그러면 $\mathcal{L}$은 λ-system이며, 위에서 보인 대로 $\mathcal{P} \subseteq \mathcal{L}$이다.

(5) Dynkin의 π-λ 정리에 의해 $\mathcal{F}_\infty = \sigma(\mathcal{P}) \subseteq \mathcal{L}$. 즉, A는 $\mathcal{F}_\infty$와 독립이다.

(6) 그런데 $A \in \mathcal{T} \subseteq \mathcal{F}_\infty$이므로, A는 자기 자신과 독립이다: $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$.

(7) 따라서 $\mathbb{P}(A) \in \{0, 1\}$.

□

Remark 15.9 (의미). 독립인 수열에서 "무한히 멀리"의 정보만으로 결정되는 사건은 확률이 0 또는 1로 퇴화한다. 예를 들어 i.i.d. 수열의 부분합이 수렴할 확률, $\limsup$의 값 등은 0 또는 1이다.

16 L^p-space

16.1 정의와 기본 성질

Definition 16.1 (L^p 공간과 노름). 확률공간 $(\Omega, \mathcal{F}, \mathbb{P})$에서 $p \geq 1$에 대해

$$L^p := L^p(\Omega, \mathcal{F}, \mathbb{P}) := \{X : \mathbb{E}[|X|^p] < \infty\}$$

로 정의하고, (동치류 위에서)

$$\|X\|_p := \left(\mathbb{E}[|X|^p]\right)^{1/p}$$

를 L^p-노름이라 한다.

Remark 16.2. 엄밀히는 L^p는 "a.s.로 같은 함수들을 동일시한 동치류"의 공간이다. 즉, $X = Y$ a.s.이면 X와 Y는 L^p에서 같은 원소로 취급한다.

16.2 Hölder와 Minkowski

Theorem 16.3 (Hölder 부등식)**.** *$\frac{1}{p} + \frac{1}{q} = 1$ $(1 \le p, q \le \infty)$이면*

$$\mathbb{E}[|XY|] \le \|X\|_p \, \|Y\|_q.$$

특히 $p = q = 2$이면 Cauchy–Schwarz 부등식이다.

구조화된 증명(표준화 + Young). **(1)** $X = 0$ 또는 $Y = 0$이면 자명하니 $\|X\|_p, \|Y\|_q > 0$이라 하자.

(2) $U := |X|/\|X\|_p$, $V := |Y|/\|Y\|_q$로 두면 $\mathbb{E}[U^p] = \mathbb{E}[V^q] = 1$이다.

(3) Young 부등식 $uv \le \frac{u^p}{p} + \frac{v^q}{q}$를 $u = U(\omega)$, $v = V(\omega)$에 적용하면

$$UV \le \frac{U^p}{p} + \frac{V^q}{q}.$$

(4) 양변에 기댓값을 취하면

$$\mathbb{E}[UV] \le \frac{1}{p}\mathbb{E}[U^p] + \frac{1}{q}\mathbb{E}[V^q] = \frac{1}{p} + \frac{1}{q} = 1.$$

(5) 따라서 $\mathbb{E}[|XY|] = \|X\|_p \|Y\|_q \mathbb{E}[UV] \le \|X\|_p \|Y\|_q$.

□

Theorem 16.4 (Minkowski 부등식)**.** *$p \ge 1$이면*

$$\|X + Y\|_p \le \|X\|_p + \|Y\|_p.$$

구조화된 증명(Hölder 사용). **(1)** $p = 1$은 삼각부등식 $\mathbb{E}[|X + Y|] \le \mathbb{E}[|X|] + \mathbb{E}[|Y|]$로 자명하다.

(2) $p > 1$이라 하자. $|X + Y|^p = |X + Y| \cdot |X + Y|^{p-1}$이므로

$$\mathbb{E}[|X + Y|^p] \le \mathbb{E}[|X| \cdot |X + Y|^{p-1}] + \mathbb{E}[|Y| \cdot |X + Y|^{p-1}].$$

(3) Hölder 부등식(Theorem 16.3)을 $(p, q) = (p, \frac{p}{p-1})$에 적용하면

$$\mathbb{E}[|X| \cdot |X + Y|^{p-1}] \le \|X\|_p \, \||X + Y|^{p-1}\|_q = \|X\|_p \, \|X + Y\|_p^{p-1},$$

같이 $\mathbb{E}[|Y| \cdot |X + Y|^{p-1}] \le \|Y\|_p \, \|X + Y\|_p^{p-1}$.

(4) 따라서 $\mathbb{E}[|X + Y|^p] \le (\|X\|_p + \|Y\|_p)\|X + Y\|_p^{p-1}$.

(5) 양변을 $\|X + Y\|_p^{p-1}$로 나누어 결론을 얻는다. (만약 $\|X + Y\|_p = 0$이면 자명.)

□

16.3 포함관계와 완비성

Theorem 16.5 (L^p 포함관계(확률공간))**.** *확률공간에서 $1 \le r < s < \infty$이면 $L^s \subseteq L^r$이고*

$$\|X\|_r \le \|X\|_s.$$

구조화된 증명(Hölder). **(1)** $X \in L^s$라 하고 $q := \frac{s}{s-r} > 1$로 두면 $\frac{r}{s} + \frac{1}{q} = 1$이다.

(2) Hölder를 $|X|^r$과 1에 적용하면

$$\mathbb{E}[|X|^r] = \mathbb{E}[|X|^r \cdot 1] \le \left(\mathbb{E}[(|X|^r)^{s/r}]\right)^{r/s} \left(\mathbb{E}[1^q]\right)^{1/q} = (\mathbb{E}[|X|^s])^{r/s}.$$

(3) 양변에 $1/r$제곱을 취하면 $\|X\|_r \le \|X\|_s$이고, 따라서 $X \in L^r$.

□

Theorem 16.6 (L^p의 완비성(요지))**.** *$1 \le p \le \infty$에서 $(L^p, \|\cdot\|_p)$는 Banach 공간이며, 특히 $p = 2$일 때 내적*

$$\langle X, Y \rangle := \mathbb{E}[XY]$$

로 Hilbert 공간이 된다.

Remark 16.7. 완비성의 증명은 (i) a.s. 수렴하는 부분수열 추출, (ii) Fatou/DCT를 이용한 노름 수렴 확인으로 이뤄진다. 이 강의노트에서는 이후 L^2 정사영 해석에 필요한 사실로 사용한다.

17 Random vector

17.1 정의와 분포

Definition 17.1 (랜덤벡터)**.** $X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$가 $(\Omega, \mathcal{F})$에서 $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$로 가는 가측함수이면 X를 *랜덤벡터*라 한다.

Theorem 17.2 (직사각형 생성계)**.** *$\mathcal{B}(\mathbb{R}^d)$는 직사각형들의 π-system*

$$\mathcal{R} := \left\{ \prod_{i=1}^{d} (a_i, b_i] : a_i < b_i \right\}$$

가 생성한다: $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{R})$.

Remark 17.3. 따라서 $(X_1, \ldots, X_d)$의 가측성이나 독립성은 좌표별 구간 사건들로 충분히 검사할 수 있다. (Dynkin의 π-λ 트릭의 전형적 응용)

Definition 17.4 (결합분포와 주변분포)**.** 랜덤벡터 $X = (X_1, \ldots, X_d)$의 *결합분포*는 push-forward 측도

$$\mu_X := \mathbb{P} \circ X^{-1} \quad \text{on } \mathcal{B}(\mathbb{R}^d)$$

이다. 각 성분 X_i의 분포는 μ_{X_i}로 표기한다(주변분포).

17.2 공분산 행렬

Definition 17.5 (공분산 행렬)**.** $X \in L^2(\Omega; \mathbb{R}^d)$이면 공분산 행렬을

$$\Sigma_X := \Big(\mathrm{Cov}(X_i, X_j) \Big)_{1 \le i,j \le d}$$

로 정의한다.

Theorem 17.6 (공분산 행렬의 성질)**.** *Σ_X는 대칭이고 양의 준정부호(positive semidefinite)이다. 즉, 모든 $a \in \mathbb{R}^d$에 대해 $a^\top \Sigma_X a \ge 0$.*

구조화된 증명 **(1)** 대칭성은 $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$에서 자명하다.

(2) 임의의 $a \in \mathbb{R}^d$에 대해 $Y := a^\top (X - \mathbb{E}[X]) = \sum_i a_i (X_i - \mathbb{E}[X_i])$라 두면 $Y \in L^2$.

(3) 분산의 정의로부터

$$\mathrm{Var}(Y) = \mathbb{E}[Y^2] = a^\top \Sigma_X a \ge 0.$$

□

17.3 상관계수

Definition 17.7 (상관계수)**.** $X, Y \in L^2$이고 $\mathrm{Var}(X), \mathrm{Var}(Y) > 0$이면

$$\rho_{XY} := \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} \in [-1, 1]$$

를 *상관계수*라 한다.

Theorem 17.8 (상관계수의 범위와 등호조건)**.** *$|\rho_{XY}| \le 1$이고, $|\rho_{XY}| = 1$은 $Y = aX + b$ a.s.인 경우와 동치이다.*

구조화된 증명 **(1)** Cauchy–Schwarz로부터

$$|\mathrm{Cov}(X, Y)| = |\langle X - \mathbb{E}[X],\, Y - \mathbb{E}[Y] \rangle| \le \|X - \mathbb{E}[X]\|_2 \, \|Y - \mathbb{E}[Y]\|_2 = \sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}.$$

(2) 따라서 $|\rho_{XY}| \le 1$.

(3) 등호조건은 Cauchy–Schwarz의 등호조건과 동일하며, $Y - \mathbb{E}[Y] = c(X - \mathbb{E}[X])$ a.s.와 동치이다. 이는 $Y = aX + b$ a.s. 형태와 같다.

□

Remark 17.9 (선형변환). $Y = AX + b$ $(A \in \mathbb{R}^{m\times d})$이면

$$\mathbb{E}[Y] = A\mathbb{E}[X] + b, \qquad \Sigma_Y = A\Sigma_X A^\top.$$

이는 고차원 정규분포, 선형회귀, 주성분분석(PCA)에서 기본 공식으로 반복 사용된다.

18 Gaussian vectors

18.1 정의

Definition 18.1 (가우시안(정규) 랜덤벡터)**.** 랜덤벡터 $X \in \mathbb{R}^d$가 *가우시안*이라 함은 임의의 $a \in \mathbb{R}^d$에 대해 $a^\top X$가 (1차원) 정규분포를 따르는 것이다.

Definition 18.2 (모수 표현)**.** 평균 $m \in \mathbb{R}^d$와 공분산 행렬 Σ를 갖는 가우시안 벡터를

$$X \sim \mathcal{N}(m, \Sigma)$$

로 쓴다. 이때 특성함수는

$$\varphi_X(t) = \exp\left(it^\top m - \frac{1}{2}t^\top \Sigma t\right), \qquad t \in \mathbb{R}^d.$$

18.2 선형변환과 독립성

Theorem 18.3 (선형변환의 안정성)**.** *$X \sim \mathcal{N}(m, \Sigma)$이고 $Y = AX + b$이면*

$$Y \sim \mathcal{N}(Am + b,\ A\Sigma A^\top).$$

구조화된 증명(특성함수). **(1)** 임의의 $t \in \mathbb{R}^m$에 대해 $t^\top Y = t^\top(AX + b) = (A^\top t)^\top X + t^\top b$.

(2) 특성함수 정의로

$$\varphi_Y(t) = \mathbb{E}[e^{it^\top Y}] = e^{it^\top b}\mathbb{E}[e^{i(A^\top t)^\top X}] = e^{it^\top b}\varphi_X(A^\top t).$$

(3) φ_X의 닫힌형을 대입하면

$$\varphi_Y(t) = \exp\left(it^\top(Am + b) - \frac{1}{2}t^\top(A\Sigma A^\top)t\right),$$

이는 $\mathcal{N}(Am + b, A\Sigma A^\top)$의 특성함수이므로 결론.

□

Theorem 18.4 (가우시안에서 비상관 ⇔ 독립)**.** *$X = (X_1, \dots, X_d) \sim \mathcal{N}(m, \Sigma)$이면*

$$\Sigma \text{ 가 대각행렬} \iff X_1, \dots, X_d \text{ 가 서로 독립}.$$

더 일반적으로, $(X_1, \dots, X_k)$와 $(X_{k+1}, \dots, X_d)$의 공분산 블록이 0이면 두 블록은 독립이다.

구조화된 증명(특성함수의 곱분해). **(1)** X의 특성함수는 $\varphi_X(t) = \exp(it^\top m - \frac{1}{2}t^\top \Sigma t)$.

(2) Σ가 대각행렬이면 $t^\top \Sigma t = \sum_{j=1}^d \Sigma_{jj} t_j^2$로 분해되므로

$$\varphi_X(t) = \prod_{j=1}^d \exp\left(im_j t_j - \frac{1}{2}\Sigma_{jj} t_j^2\right) = \prod_{j=1}^d \varphi_{X_j}(t_j).$$

(3) 특성함수의 곱분해는 결합분포가 곱측도(독립)임을 의미한다(유일성 정리). 따라서 $X_1, \ldots, X_d$는 독립이다.

(4) 반대로 X_j들이 독립이면 결합 특성함수는 곱으로 분해되고, 가우시안 특성함수 형태와 비교하면 교차항 $t_i t_j$의 계수가 0이어야 하므로 Σ의 비대각 성분이 0이다.

□

Remark 18.5 (응용). 가우시안에서 "비상관 = 독립"은 매우 특수하지만 강력하다. 선형회귀의 정규오차 가정, 칼만필터, 가우시안 프로세스 등에서 공분산 구조만으로 독립성을 판단할 수 있게 해 준다.

19 Convergences (확률변수의 수렴)

확률론에서 가장 중요한 두 가지 정리를 꼽으라면 단연 큰 수의 법칙**(Law of Large Numbers, LLN)**과 중심극한정리**(Central Limit Theorem, CLT)**이다. 본 장에서는 이 두 정리의 수학적 뼈대가 되는 확률변수열의 '수렴'을 엄밀한 측도론적 관점에서 재조명하고, 각 수렴의 의미와 그들 간의 계층적 포함 관계를 다룬다.

19.1 수렴의 네 가지 정의

해석학에서의 점별 수렴이나 균등 수렴 개념을 불확실성이 내재된 확률공간에 적용하기 위해, 목적에 맞추어 4가지 수렴을 정의한다.

Definition 19.1 (수렴의 종류)**.** 확률공간 $(\Omega, \mathcal{F}, \mathbb{P})$ 상의 확률변수열 $(X_n)_{n\in\mathbb{N}}$과 확률변수 X에 대하여 다음을 정의한다.

(1) Almost sure convergence (거의 확실한 수렴**)**:

$$X_n \xrightarrow{a.s.} X \iff \mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

(2) Convergence in probability (확률 수렴**)**:

$$X_n \xrightarrow{P} X \iff \forall \varepsilon > 0,\ \lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

(3) L^p **convergence (**L^p 수렴**)** (단, $p \geq 1$):

$$X_n \xrightarrow{L^p} X \iff \lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0.$$

(4) Convergence in distribution (분포 수렴 / 약수렴**)**: F_X가 X의 누적분포함수(CDF)일 때,

$$X_n \xrightarrow{d} X \iff \lim_{n\to\infty} F_{X_n}(x) = F_X(x) \text{ (모든 } F_X\text{의 연속점 } x\text{에서)}.$$

Remark 19.2 (이론의 의미와 적용).

- **a.s.** 수렴은 시간 n이 지남에 따라 개별 표본의 '경로(path)' 자체가 극한 함수에 거의 완벽하게 달라붙음을 의미한다. (강한 큰 수의 법칙)
- 확률 수렴은 표본의 크기 n이 커질 때 오차가 클 확률이 점차 사라진다는 뜻이다. 통계학에서 추정량의 일치성**(Consistency)**을 증명할 때 핵심적으로 사용된다.
- L^p 수렴 중 L^2 수렴은 오차의 분산이 0으로 감을 뜻하므로, 기계학습 및 신호처리에서 평균제곱오차(MSE)의 수렴 분석에 널리 활용된다.
- 분포 수렴은 확률변수의 값 자체가 아니라 '확률의 형태'가 수렴함을 의미한다. 추정량의 점근적 신뢰구간을 형성하는 근거가 된다. (중심극한정리)

19.2 수렴들 간의 계층적 포함관계

Theorem 19.3. *확률변수의 수렴에 대하여 다음의 포함관계가 성립한다:*

(1) $X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{P} X$.

(2) $X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{P} X$.

(3) $X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{L^r} X \quad (p \geq r \geq 1)$.

(4) $X_n \xrightarrow{P} X \implies X_n \xrightarrow{d} X$.

역방향은 일반적으로 성립하지 않으나, 상수 c에 대하여 $X_n \xrightarrow{d} c$인 경우에는 예외적으로 $X_n \xrightarrow{P} c$가 성립한다.

구조화된 증명: L^p 수렴 ⇒ 확률 수렴 가정**.** $X_n \xrightarrow{L^p} X$, 즉 $\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0$이다.
목표**.** 임의의 $\varepsilon > 0$에 대하여 $\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$임을 보이자.

(1) 임의의 $\varepsilon > 0$을 고정한다.

(2) 사건 $\{|X_n - X| \geq \varepsilon\}$는 부등식의 양변을 p제곱한 사건 $\{|X_n - X|^p \geq \varepsilon^p\}$와 동일하다.

(3) 음이 아닌 확률변수 $|X_n - X|^p$에 대하여 마르코프 부등식(Markov's inequality)을 적용한다.

  (3.1) $\mathbb{P}(|X_n - X|^p \geq \varepsilon^p) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\varepsilon^p}$.

(4) 가정에 의해 분자 $\mathbb{E}[|X_n - X|^p]$는 $n \to \infty$일 때 0으로 수렴한다.

(5) 조임 정리(Squeeze theorem)에 의해 좌변의 확률 또한 0으로 수렴한다. 즉, $X_n \xrightarrow{P} X$이다.

□

19.3 Borel-Cantelli 보조정리와 부분수열

Theorem 19.4 (Borel-Cantelli Lemma)**.** *사건열 $(A_n)_{n\in\mathbb{N}}$에 대하여 $\limsup_{n\to\infty} A_n = \{A_n \text{ i.o.}\}$ (무한히 자주 일어남)이라 하자.*

(1) (제1 보조정리) $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \implies \mathbb{P}(A_n \text{ i.o.}) = 0$.

(2) (제2 보조정리) *(A_n)이 서로 독립이고* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \implies \mathbb{P}(A_n \text{ i.o.}) = 1$.

확률 수렴은 a.s. 수렴을 직접적으로 보장하지 않지만, Borel-Cantelli 보조정리를 활용하면 *매우 빠르게 수렴하는 부분수열*을 잡아내어 a.s. 수렴을 유도할 수 있다.

Theorem 19.5. *$X_n \xrightarrow{P} X$이면, $X_{n_k} \xrightarrow{a.s.} X$를 만족하는 부분수열 $(X_{n_k})_{k\in\mathbb{N}}$가 존재한다.*

이 정리는 해석학에서 점별 수렴의 성질(예: 연속함수의 보존)을 확률 수렴에 적용하고자 할 때 필수적인 연결 고리가 된다.

19.4 큰 수의 법칙 (Law of Large Numbers)

동전을 무수히 많이 던질 때 앞면이 나올 비율이 0.5에 한없이 가까워진다는 통계적 직관을 완벽하게 수식화한 정리이다.

Theorem 19.6 (Weak LLN & Strong LLN)**.** *$(X_n)_{n\in\mathbb{N}}$이 i.i.d.이고 $\mathbb{E}[|X_1|] < \infty$라 하자. 모평균을 $\mu = \mathbb{E}[X_1]$라 하면, 표본평균 $S_n/n = \frac{1}{n}\sum_{i=1}^{n} X_i$에 대하여 다음이 성립한다.*

$$\frac{S_n}{n} \xrightarrow{P} \mu \quad (\textit{Weak LLN, 확률 수렴}),$$
$$\frac{S_n}{n} \xrightarrow{a.s.} \mu \quad (\textit{Strong LLN, 거의 확실한 수렴}).$$

구조화된 증명: 분산이 존재하는 경우의 약한 큰 수의 법칙 가정**.** (X_n)은 i.i.d.이고 $\mathbb{E}[X_1] = \mu$, $\mathrm{Var}(X_1) = \sigma^2 < \infty$이다.
목표**.** $S_n/n \xrightarrow{P} \mu$임을 보이자.

(1) 표본평균의 기댓값과 분산을 계산한다.

(1.1) 기댓값의 선형성에 의해 $\mathbb{E}[S_n/n] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \mu$.

(1.2) 확률변수들이 독립이므로 분산은 합산된다: $\mathrm{Var}(S_n/n) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{\sigma^2}{n}$.

(2) 임의의 $\varepsilon > 0$에 대하여 Chebyshev 부등식을 적용한다.

(2.1) $\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(S_n/n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$.

(3) $n \to \infty$일 때 분모가 무한대로 가므로 우변은 0으로 수렴한다.

(4) 확률 수렴의 정의에 의해 $\frac{S_n}{n} \xrightarrow{P} \mu$이다.

□

분산이 존재하지 않더라도 $\mathbb{E}[|X_1|] < \infty$만으로 거의 확실한 수렴이 성립하는 것이 현대 확률론의 아름다운 성취이다.

19.5 중심 극한 정리 (Central Limit Theorem)

큰 수의 법칙이 오차가 0으로 붕괴함을 보였다면, 중심극한정리는 그 오차를 $\sqrt{n}$ 배 확대하여 돋보기로 들여다볼 때 나타나는 *점근적 분포의 형태*가 정규분포임을 밝힌다.

Theorem 19.7 (Lindeberg-Lévy CLT)**.** *(X_n)이 i.i.d., $\mathbb{E}[X_1] = \mu$, $\mathrm{Var}(X_1) = \sigma^2 \in (0, \infty)$이면,*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

구조화된 증명: 특성함수를 이용한 중심극한정리 가정**.** (X_n)은 i.i.d.이고 $\mathbb{E}[X_1] = \mu$, $\mathrm{Var}(X_1) = \sigma^2 \in (0, \infty)$이다.
목표**.** $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1)$임을 Lévy의 연속성 정리를 이용해 보이자.

(1) 표준화된 확률변수 $Y_i = \frac{X_i - \mu}{\sigma}$를 정의하자. Y_i는 i.i.d.이며 $\mathbb{E}[Y_1] = 0$, $\mathbb{E}[Y_1^2] = 1$이다.

(2) Z_n을 Y_i의 합으로 다시 쓰면 $Z_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i$이다.

(3) Y_1의 특성함수 $\varphi_Y(t) = \mathbb{E}[e^{itY_1}]$의 테일러 전개를 구하자.

(3.1) 2차 모멘트가 유한하므로, 특성함수는 $t = 0$ 근방에서 2차까지 전개 가능하다.

(3.2) $\varphi_Y(0) = 1$, $\varphi_Y'(0) = i\mathbb{E}[Y_1] = 0$, $\varphi_Y''(0) = i^2\mathbb{E}[Y_1^2] = -1$이다.

(3.3) 따라서 $\varphi_Y(t) = 1 - \frac{t^2}{2} + o(t^2)$ (as $t \to 0$)이다.

(4) 독립 확률변수 합의 특성함수 성질에 의해 Z_n의 특성함수 $\varphi_{Z_n}(t)$를 계산한다.

(4.1) $\varphi_{Z_n}(t) = \mathbb{E}\left[\exp\left(it\frac{1}{\sqrt{n}}\sum_{j=1}^n Y_j\right)\right] = \prod_{j=1}^n \mathbb{E}\left[\exp\left(i\frac{t}{\sqrt{n}}Y_j\right)\right]$.

(4.2) 즉, $\varphi_{Z_n}(t) = \left[\varphi_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n$이다.

(5) 점근 전개를 대입하고 극한을 취한다.

(5.1) 고정된 t에 대해 $n \to \infty$이면 $t/\sqrt{n} \to 0$이다.

(5.2) $\varphi_{Z_n}(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n$.

(5.3) 해석학의 표준 극한 $\lim_{n\to\infty}(1 + \frac{x_n}{n})^n = e^x$ (단, $x_n \to x$)를 적용하면 $\lim_{n\to\infty}\varphi_{Z_n}(t) = e^{-t^2/2}$이다.

(6) Lévy 연속성 정리에 의해 $Z_n \xrightarrow{d} \mathcal{N}(0, 1)$이다.

□

19.6 분포 수렴의 추가 도구들

실제 통계적 추론에서는 X_n 뿐만 아니라 그 연속 함수 변환이나 오차항이 포함된 수식을 다루게 되므로 다음 도구들이 빈번하게 쓰인다.

Theorem 19.8 (Continuous Mapping Theorem)**.** *$X_n \xrightarrow{d} X$ (또는 P, a.s. 수렴)이고 g가 연속함수이면, $g(X_n) \xrightarrow{d} g(X)$이다.*

Theorem 19.9 (Slutsky's Theorem)**.** *$X_n \xrightarrow{d} X$이고 $Y_n \xrightarrow{P} c$ (상수)이면,*

$$X_n + Y_n \xrightarrow{d} X + c, \quad X_n Y_n \xrightarrow{d} cX, \quad X_n/Y_n \xrightarrow{d} X/c \ (c \neq 0).$$

Theorem 19.10 (Delta Method)**.** *$\sqrt{n}(X_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$이고 g가 μ에서 미분 가능하며 $g'(\mu) \neq 0$ 이면,*

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{d} \mathcal{N}(0, \sigma^2 [g'(\mu)]^2).$$

Delta method는 Taylor 전개의 1차 근사를 확률 공간으로 가져온 것으로, 통계학에서 비선형 추정량의 점근적 분산(Asymptotic variance)과 신뢰구간을 유도할 때 막강한 위력을 발휘한다.

20 Conditional Expectation (조건부 기댓값)

기댓값이 상수를 반환하는 결정론적 연산이라면, 조건부 기댓값은 주어진 정보(부분 σ-algebra)에 따라 가장 합리적인 추정치를 반환하는 *확률변수*이다. 본 장에서는 측도론적 조건부 기댓값의 정의와 성질을 엄밀히 다룬다.

20.1 측도론적 정의 (Radon-Nikodym 정리 기반)

연속 확률변수 X에 대해 $X = x$일 확률은 0이므로 학부 수준의 분수식 $\mathbb{P}(A \cap B)/\mathbb{P}(B)$로는 조건부 기댓값을 온전히 정의할 수 없다.

Definition 20.1 (조건부 기댓값)**.** 확률공간 $(\Omega, \mathcal{F}, \mathbb{P})$, 적분가능한 확률변수 X, 부분 시그마 대수 $\mathcal{G} \subseteq \mathcal{F}$에 대해, X의 $\mathcal{G}$에 대한 조건부 기댓값 $\mathbb{E}[X|\mathcal{G}]$는 다음 두 조건을 만족하는 (a.s.) 유일한 확률변수이다:

(1) 가측성 **(Measurability)**: $\mathbb{E}[X|\mathcal{G}]$는 $\mathcal{G}$-measurable이다. (즉, 주어진 정보 $\mathcal{G}$만으로 그 값을 결정할 수 있다.)

(2) 부분 적분 일치성 **(Partial averaging)**: 임의의 사건 $A \in \mathcal{G}$에 대하여, $\int_A \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P} = \int_A X\, d\mathbb{P}$.

이 정의는 측도론의 Radon-Nikodym 정리를 활용하여 존재성과 유일성을 수학적으로 완벽하게 보장받는다. (Theorem 12.6 참조).

20.2 조건부 기댓값의 핵심 성질

조건부 기댓값은 선형성, 단조성 등 일반적인 기댓값의 훌륭한 성질을 모두 상속받는다.

Theorem 20.2. **(1) (Taking out what is known)** *Y가 $\mathcal{G}$-measurable이고 유계이면, $\mathbb{E}[XY|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}]$. (이미 알고 있는 정보는 상수로 취급하여 기댓값 기호 밖으로 뺄 수 있다.)*

(2) (독립성 분리**)** *X가 $\mathcal{G}$와 독립이면, $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.*

(3) (전기댓값의 법칙**)** *$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.*

마팅게일 이론을 전개하는 데 있어 절대 빠질 수 없는 성질이 바로 Tower property이다. "거친 관측(작은 σ-algebra)이 미세한 관측(큰 σ-algebra)을 덮어쓴다"는 직관을 갖는다.

구조화된 증명: Tower property. 가정**.** $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$이고 X는 적분 가능하다.
목표**.** 확률변수 $Z = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$가 X의 $\mathcal{H}$에 대한 조건부 기댓값 정의를 만족함을 보이자.

(1) **(가측성)** Z는 $\mathcal{H}$에 대한 조건부 기댓값이므로 정의에 의해 이미 $\mathcal{H}$-measurable이다.

(2) **(부분 적분 일치성)** 임의의 가측집합 $A \in \mathcal{H}$를 고정하자.

(2.1) Z의 정의에 의해, $\int_A Z\, d\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P}$이다.

(2.2) $\mathcal{H} \subseteq \mathcal{G}$이므로, 선택한 A는 $A \in \mathcal{G}$도 만족한다.

(2.3) 안쪽의 조건부 기댓값 $\mathbb{E}[X|\mathcal{G}]$의 정의에 의해, $A \in \mathcal{G}$이므로 $\int_A \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P} = \int_A X\, d\mathbb{P}$가 성립한다.

(3) 두 적분 식을 연결하면 $\int_A Z\, d\mathbb{P} = \int_A X\, d\mathbb{P}$가 성립한다.

(4) Z는 $\mathcal{H}$-measurable이며, 모든 $A \in \mathcal{H}$에 대해 X와 부분 적분이 일치하므로 Radon-Nikodym 유일성에 의해 $Z = \mathbb{E}[X|\mathcal{H}]$ (a.s.)이다.

□

20.3 조건부 분산

Definition 20.3. $V(X|\mathcal{G}) := \mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])^2|\mathcal{G}] = \mathbb{E}[X^2|\mathcal{G}] - (\mathbb{E}[X|\mathcal{G}])^2$.

Theorem 20.4 (전분산 공식, Law of total variance)**.**

$$V(X) = \mathbb{E}[V(X|\mathcal{G})] + V(\mathbb{E}[X|\mathcal{G}]).$$

이 공식은 "전체 분산 = 그룹 내 분산의 기댓값 + 그룹 간 평균의 분산"으로 해석되며, 통계학의 ANOVA (분산분석) 및 결정 트리(Decision Tree)의 불순도 감소를 설명하는 이론적 토대가 된다.

21 Orthogonal projection and L^2 space

21.1 정사영 정리 (Projection Theorem) 복습

Theorem 21.1. *내적공간 $(V, \langle\cdot,\cdot\rangle)$의 닫힌(closed) 부분공간 $W \subseteq V$와 벡터 $x \in V$가 주어졌을 때, 다음을 만족하는 $u \in W$가 유일하게 존재한다: (1) $x - u \in W^\perp$ (오차 벡터가 W의 모든 원소와 직교한다). (2) $\|x - u\| = \inf_{w \in W} \|x - w\|$ (최단 거리). 이 u를 x의 W 위로의* 정사영*이라 한다.*

22 Conditional expectation as Projection on L^2

우리가 예측하고 싶은 '타겟'을 Y, 현재 가지고 있는 '특성 정보'를 $\mathcal{G}$라고 하자. $\mathcal{G}$만으로 알 수 있는 유한 분산 확률변수들의 집합 $L^2(\Omega, \mathcal{G}, \mathbb{P})$는 $L^2(\Omega, \mathcal{F}, \mathbb{P})$ 힐베르트 공간의 닫힌 부분공간이다. 우리는 조건부 기댓값을 이 힐베르트 공간에서의 기하학적 메타포로 명쾌하게 이해할 수 있다.

Theorem 22.1 (최적 예측자로서의 조건부 기댓값)**.** *$Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$이고, $\mathcal{G} \subseteq \mathcal{F}$이다. 조건부 기댓값 $Y^* = \mathbb{E}[Y|\mathcal{G}]$는 정보 $\mathcal{G}$ 하에서 평균제곱오차(MSE)를 최소화하는 유일한 예측 함수이다.*

$$Y^* = \underset{Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})}{\arg\min}\ \mathbb{E}[(Y - Z)^2].$$

구조화된 증명: 직교성 원리 및 피타고라스 정리 가정**.** $Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$이고 $Y^* = \mathbb{E}[Y|\mathcal{G}]$이다.
목표**.** 임의의 $Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$에 대해 $\mathbb{E}[(Y - Z)^2] \geq \mathbb{E}[(Y - Y^*)^2]$임을 보이자.

(1) 오차 제곱을 Y^*를 기준으로 편차 분해한다.

(1.1) $\mathbb{E}[(Y - Z)^2] = \mathbb{E}[((Y - Y^*) + (Y^* - Z))^2]$.

(1.2) 전개하면 $\mathbb{E}[(Y - Y^*)^2] + 2\mathbb{E}[(Y - Y^*)(Y^* - Z)] + \mathbb{E}[(Y^* - Z)^2]$이다.

(2) 교차항(Cross-term) $\mathbb{E}[(Y - Y^*)(Y^* - Z)]$가 0임을 직교성으로 증명하자.

(2.1) $W := Y^* - Z$라 하면, Y^*와 Z 모두 $\mathcal{G}$-measurable이므로 W도 $\mathcal{G}$-measurable이다.

(2.2) 전 기댓값의 법칙(Tower property)에 의해 교차항은 $\mathbb{E}[\mathbb{E}[(Y - Y^*)W|\mathcal{G}]]$와 같다.

(2.3) W가 $\mathcal{G}$-measurable이므로 밖으로 뺄 수 있다: $\mathbb{E}[W\mathbb{E}[(Y - Y^*)|\mathcal{G}]]$.

(2.4) 조건부 기댓값의 선형성에 의해 $\mathbb{E}[(Y - Y^*)|\mathcal{G}] = \mathbb{E}[Y|\mathcal{G}] - \mathbb{E}[Y^*|\mathcal{G}] = Y^* - Y^* = 0$이다.

(2.5) 따라서 교차항의 기댓값은 0이 된다.

(3) 피타고라스 정리가 성립한다: $\mathbb{E}[(Y - Z)^2] = \mathbb{E}[(Y - Y^*)^2] + \mathbb{E}[(Y^* - Z)^2]$.

(4) $\mathbb{E}[(Y^* - Z)^2] \geq 0$이므로 항상 $\mathbb{E}[(Y - Z)^2] \geq \mathbb{E}[(Y - Y^*)^2]$이다. 이 식은 오직 $Z = Y^*$ (a.s.)일 때만 최솟값을 갖는다.

□

Remark 22.2 (기계학습과의 연결). 기계학습의 선형 회귀(Linear Regression)나 신경망이 MSE 손실 함수를 최소화하도록 학습될 때, 모델이 한없이 근사하려고 애쓰는 궁극적인 정답 함수(Target function)가 다름 아닌 고차원 데이터 공간에서의 조건부 기댓값 $\mathbb{E}[Y|X]$이다. 또한 신호 처리의 칼만 필터(Kalman Filter)는 매 시점마다 새로운 관측 정보의 부분공간으로 상태 변수를 정사영(Projection)하는 재귀적 과정과 정확히 일치한다.

23 Martingale (마팅게일) 입문

마팅게일은 조건부 기댓값을 시간에 따른 확률과정(Stochastic process)으로 확장한 것이다. "지금까지의 모든 정보를 고려했을 때, 미래의 기댓값은 정확히 현재의 값과 같다"는 "공평한 게임(fair game)" 철학을 수학적으로 담고 있다.

23.1 Filtration과 Adapted process

시간의 흐름에 따라 관측 정보가 축적되는 과정을 시그마 대수의 단조 증가열로 추상화한다.

Definition 23.1 (Filtration과 Adapted)**.** 확률공간의 증가하는 σ-algebra 수열 $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}$를 *Filtration*(여과)이라 한다.
확률과정 $(X_n)_{n\geq 0}$이 모든 n에 대해 X_n이 $\mathcal{F}_n$-measurable이면, (X_n)은 $(\mathcal{F}_n)$에 *adapted*(적응)되었다고 한다. 이는 의사결정에서 미래를 미리 엿볼 수 없음을 뜻한다.

23.2 마팅게일의 정의

Definition 23.2 (Martingale)**.** Adapted이고 integrable한($\mathbb{E}[|M_n|] < \infty$) 확률과정 $(M_n)_{n\geq 0}$이 다음을 만족하면 *마팅게일*이라 한다:

$$\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n \quad (a.s.) \quad \forall n \geq 0.$$

(위 등식을 $\leq$로 바꾸면 Supermartingale(참여자에게 불리), $\geq$로 바꾸면 Submartingale(참여자에게 유리)이라 한다.)

마팅게일의 가장 근본적인 성질은 Tower Property를 반복 적용하여 유도된다. 즉, 모든 $m > n$에 대해 $\mathbb{E}[M_m|\mathcal{F}_n] = M_n$ 이며, 임의의 시점 n에 대해 무조건 $\mathbb{E}[M_n] = \mathbb{E}[M_0]$이다.

23.3 마팅게일 수렴 정리와 Optional Stopping

Theorem 23.3 (Doob's Martingale Convergence Theorem)**.** *(M_n)이 supermartingale이고 하방으로 유계 ($\sup_n \mathbb{E}[M_n^-] < \infty$)이면, 확률 1로 $\lim_{n\to\infty} M_n = M_\infty$가 존재하며 그 극한값은 유한($M_\infty \in L^1$)하다.*

그렇다면 도박사가 "적절한 타이밍(정지 시간, Stopping time τ)"에 게임을 멈추면 평균적으로 이득을 볼 수 있을까?

Definition 23.4 (Stopping Time)**.** 확률변수 $\tau : \Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$가 모든 n에 대해 $\{\tau \leq n\} \in \mathcal{F}_n$을 만족하면 stopping time이라 한다.

Theorem 23.5 (Optional Stopping Theorem, OST). *(M_n)이 마팅게일이고 τ가 유한한(a.s. $\tau < \infty$) 정지 시간이라 하자. τ가 유계(bounded)이거나 적절한 증분 유계 조건을 만족하면, 멈춘 시점의 자산 기댓값은 초기 자본과 동일하게 보존된다:*

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_0].$$

구조화된 증명: τ가 유계인 경우의 OST. 가정. (M_n)은 마팅게일이고, 정지 시간 τ가 어떤 상수 N에 대해 $\tau \le N$ a.s.를 만족한다.
목표. $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$임을 텔레스코핑 합을 통해 증명하자.

(1) $M_\tau - M_0$를 차분(difference)의 합으로 전개한다.

- **(1.1)** $M_\tau - M_0 = \sum_{k=1}^{\tau}(M_k - M_{k-1})$.
- **(1.2)** $\tau \le N$이므로 지시함수를 도입해 합의 범위를 N까지로 고정한다: $\sum_{k=1}^{N}(M_k - M_{k-1})\mathbf{1}_{\{\tau\ge k\}}$.

(2) 양변에 기댓값을 취하고 선형성에 의해 합을 분리한다: $\mathbb{E}[M_\tau - M_0] = \sum_{k=1}^{N}\mathbb{E}[(M_k - M_{k-1})\mathbf{1}_{\{\tau\ge k\}}]$.

(3) 각 항이 0임을 증명한다.

- **(3.1)** 사건 $\{\tau \ge k\}$는 $\{\tau \le k-1\}^c$와 같다. τ가 정지 시간이므로 $\{\tau \le k-1\} \in \mathcal{F}_{k-1}$이다. 즉, $\mathbf{1}_{\{\tau\ge k\}}$는 $\mathcal{F}_{k-1}$-measurable이다.
- **(3.2)** Tower property에 의해 $\mathbb{E}[(M_k - M_{k-1})\mathbf{1}_{\{\tau\ge k\}}] = \mathbb{E}\big[\mathbb{E}[(M_k - M_{k-1})\mathbf{1}_{\{\tau\ge k\}} \mid \mathcal{F}_{k-1}]\big]$.
- **(3.3)** Taking out 성질을 적용하면 $\mathbb{E}\big[\mathbf{1}_{\{\tau\ge k\}}\mathbb{E}[M_k - M_{k-1} \mid \mathcal{F}_{k-1}]\big]$이다.
- **(3.4)** 마팅게일 정의에 의해 $\mathbb{E}[M_k - M_{k-1} \mid \mathcal{F}_{k-1}] = 0$이므로 기댓값은 0이다.

(4) 모든 항이 0이므로 $\mathbb{E}[M_\tau - M_0] = 0$이다. 따라서 $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$.

□

Remark 23.6 (금융 수학으로의 응용). OST는 현대 금융 수학(블랙-숄즈 모형 등)의 중심 기둥이다. "위험중립 측도(Risk-neutral measure) 하에서 할인된 자산 가격 과정은 마팅게일이다"라는 사실과 결부되어, 시장에 무위험 차익거래(Arbitrage)가 존재하지 않는다는 강력하고 우아한 수학적 선언이 된다.

23.4 Doob's Maximal Inequality

Theorem 23.7. *(M_n)이 submartingale이고 $\lambda > 0$이면, 확률과정의 전체 경로 중 최댓값에 대해 다음 꼬리 한계가 성립한다:*

$$\mathbb{P}\left(\max_{1\le k\le n} M_k \ge \lambda\right) \le \frac{\mathbb{E}[M_n^+]}{\lambda}.$$

24 Advanced Concentration Inequalities (고급 집중 부등식)

큰 수의 법칙(LLN)은 표본 수가 $n \to \infty$로 커질 때의 수렴을 설명한다. 그러나 고차원 통계학, 인공지능, 및 기계학습 이론(Statistical Learning Theory)에서는 "유한한 n개의 데이터만 가지고 있을 때, 오차가 ε 이상 날 확률이 최대 얼마인가?"라는 비점근적**(Non-asymptotic)** 유한 표본 한계**(Finite sample bound)**가 절실하다. 집중 부등식은 이 꼬리 확률(Tail probability)이 데이터 수에 따라 기하급수적(Exponentially)으로 감소함을 보장한다.

24.1 Chernoff Bound Framework

모든 고급 집중 부등식은 지수 함수의 단조증가성과 Markov 부등식을 절묘하게 조합한 Chernoff Bound에서 시작된다.

Theorem 24.1 (Chernoff Bound). *임의의 확률변수 X와 임의의 $\lambda > 0$에 대해,*

$$\mathbb{P}(X \ge t) = \mathbb{P}(e^{\lambda X} \ge e^{\lambda t}) \le e^{-\lambda t}\mathbb{E}[e^{\lambda X}].$$

적률생성함수(MGF) $\mathbb{E}[e^{\lambda X}]$의 상한을 분포 특성에 맞게 구한 후, 우변을 최소화하는 최적의 λ를 찾아 대입하면 강력한 부등식이 탄생한다.

24.2 Hoeffding's Inequality

확률변수들이 유계(bounded)인 상황에서 가장 흔하고 위력적으로 쓰이는 부등식이다.

Lemma 24.2 (Hoeffding's Lemma)**.** *$\mathbb{E}[X] = 0$이고 $a \le X \le b$ (a.s.)이면, 임의의 $\lambda > 0$에 대해 다음 MGF 상한이 성립한다:*

$$\mathbb{E}[e^{\lambda X}] \le \exp\left(\frac{\lambda^2(b-a)^2}{8}\right).$$

구조화된 증명: Hoeffding's Inequality. 가정**.** $X_1, \dots, X_n$은 서로 독립이고 $\mathbb{E}[X_i] = 0$, $a_i \le X_i \le b_i$ (a.s.)이다.
목표**.** $S_n = \sum_{i=1}^n X_i$에 대해 $\mathbb{P}(S_n \ge t) \le \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$를 증명하자.

(1) 부분합 S_n에 대하여 Chernoff Bound 테크닉을 적용한다.

(1.1) 임의의 $\lambda > 0$에 대해 $\mathbb{P}(S_n \ge t) = \mathbb{P}(e^{\lambda S_n} \ge e^{\lambda t})$.

(1.2) Markov 부등식에 의해 $\mathbb{P}(e^{\lambda S_n} \ge e^{\lambda t}) \le e^{-\lambda t}\mathbb{E}[e^{\lambda S_n}]$이다.

(2) X_i들이 상호 독립이므로 결합 MGF를 개별 MGF의 곱으로 분해한다.

(2.1) $\mathbb{E}[e^{\lambda S_n}] = \mathbb{E}\left[\exp\left(\sum_{i=1}^n \lambda X_i\right)\right] = \prod_{i=1}^n \mathbb{E}[e^{\lambda X_i}]$.

(3) 각 X_i에 대하여 Hoeffding's Lemma를 적용하고 하나로 묶는다.

(3.1) $\mathbb{E}[e^{\lambda X_i}] \le \exp\left(\frac{\lambda^2(b_i - a_i)^2}{8}\right)$.

(3.2) 이를 곱하면 $\prod_{i=1}^n \mathbb{E}[e^{\lambda X_i}] \le \exp\left(\frac{\lambda^2}{8}\sum_{i=1}^n (b_i - a_i)^2\right)$이다.

(4) 부등식을 종합하고 최적의 λ를 찾는다.

(4.1) $\mathbb{P}(S_n \ge t) \le \exp\left(-\lambda t + \frac{\lambda^2}{8}\sum_{i=1}^n (b_i - a_i)^2\right)$.

(4.2) 지수 부분을 λ에 대해 미분하여 최솟값을 갖는 $\lambda^* = \frac{4t}{\sum (b_i - a_i)^2}$를 얻는다.

(4.3) 원래 식에 λ^*를 대입하여 정리하면 최댓값 상한 $\exp\left(-\frac{2t^2}{\sum (b_i - a_i)^2}\right)$를 얻는다. 양방향 절대값 확률의 경우 대칭성에 의해 2를 곱한다.

□

24.3 Sub-Gaussian과 Sub-Exponential Random Variables

Hoeffding 보조정리가 지수 함수의 λ^2에 비례하는 형태임에 착안하여, 꼬리(tail)가 정규분포(가우시안)처럼 얇게 감소하는 확률변수군을 추상화한다.

Definition 24.3 (Sub-Gaussian)**.** $\mathbb{E}[X] = 0$인 확률변수 X가 분산 파라미터 σ^2의 *Sub-Gaussian*이라 함은,

$$\mathbb{E}[e^{\lambda X}] \le \exp\left(\frac{\sigma^2\lambda^2}{2}\right), \quad \forall \lambda \in \mathbb{R}$$

을 만족하는 것이다. 모든 유계(Bounded) 분포와 정규분포가 이에 속하며, 꼬리 확률이 $\exp(-t^2/2\sigma^2)$ 수준으로 극도로 빠르게 붕괴한다.

Sub-Gaussian 변수를 제곱(예: χ^2 분포)하면 꼬리가 두꺼워져 모든 λ가 아닌 원점 근방($|\lambda| \le 1/b$)에서만 MGF 한계를 갖는데, 이를 *Sub-Exponential* 분포라 부른다.

24.4 Bernstein's Inequality

Hoeffding 부등식은 변수의 유일한 한계인 '최악의 양 끝 구간 $b - a$' 정보만을 사용할 뿐, 데이터의 실제 분산(σ^2) 정보를 반영하지 않아 헐거운(loose) 경계를 제공하는 경우가 많다. Bernstein은 분산을 수식에 끌어들여 이를 극복했다.

Theorem 24.4 (Bernstein's Inequality)**.** *$(X_i)_{i=1}^n$이 독립이고 $\mathbb{E}[X_i] = 0$, 모든 변수가 유계 $|X_i| \le M$ (a.s.)를 만족한다고 하자. 표본 분산 평균을 $\sigma^2 = \frac{1}{n}\sum_{i=1}^n \mathrm{Var}(X_i)$라 할 때,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i\right| \ge t\right) \le 2\exp\left(-\frac{nt^2/2}{\sigma^2 + Mt/3}\right).$$

Remark 24.5 (의미 및 ML로의 적용)*.* 수식을 자세히 들여다보면 오차 t가 작을 때 지수 항은 $-nt^2/2\sigma^2$로 근사되어 CLT의 분산 σ^2을 완벽하게 포착한다. 만약 학습 데이터의 분산이 작다면, Bernstein 부등식은 Hoeffding 부등식보다 데이터가 중심에 훨씬 뾰족하게 집중되어 있음을 날카롭게 파악해 낸다. 이 부등식은 PAC-Learning에서 모델의 복잡도 제어와 경험적 위험 최소화(Empirical Risk Minimization) 성능을 증명할 때 결정적 역할을 수행한다.

25 맺음말

본 강의록에서는 확률론의 뼈대를 이루는 내용을 수리적 엄밀함을 잃지 않으면서도 직관적인 해석을 곁들여 다루었다.

해석학과 측도론의 언어로 확률공간을 세우고 르베그 적분이라는 정교한 체로 기댓값을 걸러냈다. 이를 바탕으로 큰 수의 법칙과 중심극한정리라는 거시적 극한의 철학을 엄밀히 증명했다. 특히 조건부 기댓값을 힐베르트 공간 L^2에서의 정사영(Orthogonal Projection)으로 바라보는 기하학적 통찰은 마팅게일(Martingale) 이론이라는 확률과정론의 꽃을 피우는 완벽한 토대가 되었으며, 비점근적 유한 표본 오차를 지수적으로 억제하는 집중 부등식(Concentration Inequality)은 현대 기계학습 및 인공지능 기초 이론의 든든한 안전망이 되어준다.

이 굳건한 기초를 바탕으로 확률미분방정식(SDE)과 이토 미적분(Ito Calculus), 고차원 통계학, 통계적 학습 이론(Statistical Learning Theory)이라는 더 넓고 깊은 바다로 훌륭하게 항해해 나갈 수 있기를 바란다.

□