44.006 — University of Seoul, Spring 2026

# Lecture 5: Empirical Process

May, 11, 2026

*Lecturer*: *Jong-June Jeon* *Scribe*: *Jong-June Jeon*

# 1 Empirical Process 이름의 유래

$$P_n m_\theta, P m_\theta.$$

왜 empirical process 라고 하였을까? 무엇인 process 인가? 이를 이해하기 위해서 우리에게 익숙한 형태로 $P_n m_\theta$와 $P m_\theta$

확률과정은 보통 시계열의 형태로 처음 등장한다. 시간 $t$에 따라 확률변수 $X_t$가 변화하는 과정 $\{X_t : t \in \mathbb{R}\}$ 또는 $\{X_t : t \in \mathbb{Z}\}$를 생각하게 된다. 하지만 확률과정의 정의 자체는 "시간"에 본질적으로 묶여 있지 않다. 확률과정이란 단지 어떤 집합 $T$ 위에서 정의된 확률변수들의 모임 $\{X_t\}_{t\in T}$일 뿐이다. $t \in \mathbb{Z}$은 확률변수에 대한 index로 볼 수 있고, $\{X_t : t \in \mathbb{Z}\}$는 index set에 대응되는 확률변수의 모임으로 볼 수 있다. 만약 $P_n m_\theta$의 index set $\Theta \subset \mathbb{R}^k$를 생각한다면 $\{P_n m_\theta : \theta \in \Theta\}$를 확률과정으로 생각할 수 있다. 앞서 우리가 $P_n m_\theta$를 경험적 확률과정 (empirical process)이라 불렀던 하나의 이유다. 그러면 자연스럽게 다음과 같은 질문이 생긴다. 전통적으로는 $T = \mathbb{R}$ 같은 1차원 실수 시간축 어떻게 다차원의 index set 으로 확장이 되었는가? 이를 알기 위해서는 시계열확률과정에서 수학/통계학자들이 밝혀낸 결과들이 확률과정읜 index set의 어떤 특성으로 부터 유도된 것인지 알아보는 것이 필요하다.

## 1.1 시계열에서 시간 실수축이 의미를 가지는 첫 번째 이유: 순서 구조

시계열에서 인덱스 공간이 실수 $\mathbb{R}$로 주어지는 이유는, 실수가 단순한 집합이 아니라 순서(total order)를 가진다는 특성으로 "과거–현재–미래"라는 시간의 방향성을 표현할 수 있기 때문이다. filtration $\{\mathcal{F}_t\}$[1] 를 정의할 때, $s \le t$이면 $\mathcal{F}_s \subset \mathcal{F}_t$가 되는 단조성은 시간축의 순서 구조를 전제로 한다. 또한 실수는 완비성(completeness)을 가지기 때문에 "시점이 가까워진다"는 개념을 자연스럽게 정의할 수 있다.

하지만 만약 우리가 다루는 확률과정에서 이러한 순서가 필요하지 않다면, 즉 "시간의 흐름" 자체가 본질이 아니라면 굳이 실수축을 인덱스로 둘 이유는 없다. $\mathbb{R}^d$위의 원소들을 다루는 데 순서가 반드시 필요하지는 않다. 순서라는 개념이 필요하지 않은 인덱스 공간위에 정의된 확률과정을 정의할 수 있을까? 그러한 확률과정을 다루는 것이 유용할까는 의문이 들겠지만 수학자들은 이미 이를 성공적으로

[1] 확률과정에서 $t$ 시점까지 누적된 정보

바로 Empirical process가 대표적인 예이다. 거기서는 “먼저 평가된 함수”와 “나중에 평가된 함수”라는 개념이 없고, 순서 대신 함수들 사이의 “가까움”이 중요해진다.

## 1.2 두 번째 이유: 과정의 존재와 측도가능성

확률과정은 단순히 각 $X_t$가 존재하는 것으로 끝나지 않는다. 우리는 전체 과정 $\{X_t\}_{t\in T}$를 하나의 무한차원 확률변수로 생각한다. 즉 $\omega$가 주어졌을 때 $\omega \mapsto (X_t(\omega))_{t\in T}$라는 경로(path)가 만들어진다. Kolmogorov extension theorem은 finite-dimensional distribution이 일관되게 주어지면 임의의 인덱스 집합 $T$에 대해서도 확률과정을 구성할 수 있음을 보장한다. 하지만 이 정리는 단지 “어떤 과정이 존재한다”는 것만 말해줄 뿐, 그 과정이 좋은 성질을 가진다는 것을 자동으로 보장하지는 않는다.
예를 들어 다음과 같은 질문들이 남는다.

- 경로가 측도가능(measurable)한가?
- 연속적인 modification이 존재하는가?
- 과정이 약수렴할 때 tightness를 논할 수 있는가?

이런 문제들을 다루기 위해서는 인덱스 집합 $T$가 단순한 집합이 아니라 위상적 구조를 가져야 한다. 인덱스 집합위에 특히 metric 구조가 있으면 “가까워진다”는 개념이 생기고, 연속성이나 극한을 정의할 수 있다.또한 separability가 있으면 가산 dense subset을 통해 측도가능성 문제가 훨씬 단순해진다 이때 가장 자연스럽게 등장하는 공간이 *Polish space*이다.

## 1.3 Polish Space의 역할

Polish space란 separable하고 complete한 metric space를 말한다. 실수 $\mathbb{R}$는 대표적인 Polish space이다. Polish 공간이 중요한 이유는 다음과 같다.

- Borel $\sigma$-algebra가 확률론적으로 잘 정의된다.
- 확률측도의 tightness와 약수렴 이론이 자연스럽게 확장된다.
- 함수공간에서의 과정 수렴을 다룰 수 있다.

즉 실수축에서 우리가 암묵적으로 사용하던 좋은 성질들은 대부분 Polish 공간으로 일반화될 수 있다. 따라서 확률과정의 인덱스 공간은 실수축이 아니더라도 Polish metric space이면 충분한 경우가 많다. 주의할 점은 Polish Metric space에서 separable하고 complete라른 성질은 어떤 metric 을 사용하는가에 의존한다. Metric을 바꾸게 되면 이전에 성립하던 확률공간에서의 성질들이 성립하지 않을 수 있다.

## 1.4 Empirical Process에서의 확장

Empirical process에서는 인덱스가 시간이 아니라 함수 클래스이다. 표본 $X_1, \ldots, X_n$이 주어졌을 때, 사실 empirical process는 앞서 우리는 $P_n m_\theta$를 empirical process 라고 불렀지만 보통은 $G_n(\theta) = \sqrt{n}(P_n - P)m_\theta$ 로 정의되고, 이는 $\theta$에 의해 인덱싱된다. 그 이유는 empirical process가 동전 던지기 실험처럼 극한분포가 존재하는 형태로 만들기 위함이다. index 를 고정했을때 랜덤 워크는 모든 시점에서 평균이 0이고 랜덤워크 항에 $\sqrt{n}$ 을 곱하여 시점을 고정하였을 때 극한 분포 수렴하도록 만들어져있다. $P_n m_\theta$의 평균을 0으로 만들고 $\theta$를 보정 하였을 때 분포수렴할 수 있도록 $\sqrt{n}$을 곱한항인 $G_n(\theta) = \sqrt{n}(P_n - P)m_\theta$를 empirical process로 정의하면 기존에 우리가 분석하였던 Wiener Process 의 유도 절차를 따를 수 있을 것이다. 따라서 경험적 확률과정은

$$\{G_n(\theta) : \theta \in \Theta\}$$

처럼 함수공간 위에서 정의된다.
한편 경험적 확률 과정을 정의하는 지표공간이 실수가 아닌 매트릭 공간이 됨을 허용 할 수 있다면, 지표공간이 다차원 실수 공간일 필요는 없다. 여기서 중요한 것은 순서가 아니라. 중요한 것은 매트릭 공간 상의 Jeong의 된 원소들 간의 metric 구조이다. 예를 들어 대표적으로 함수공간 $\mathcal{F}$ 위에 $L^2(P)$ 거리를 정의하고 $d(f, g) = \sqrt{P(f - g)^2}$로 정의한 metric 공간을 지표공간으로 확장 가능성을 생각할 수 있다.이런 경우에는 경험적 확률과정이

$$\{G_n(f) : f \in \mathcal{F}\}$$

로 정의될 것이다. 지표 공간의 특성을 통해 확률과정의 연속성, tightness, entropy 조건 등을 논하게 된다.

## 1.5 정리

시계열에서 실수축이 중요했던 이유는 순서 구조를 통해 시간의 방향성을 표현할 수 있었고, 또한 완비성과 separability를 통해 측도가능성과 수렴 이론을 전개할 수 있었기 때문이다. 하지만 empirical process처럼 순서가 필요하지 않은 경우에는 시간축을 유지할 이유가 없다. 이때 확률과정 이론은 실수축이 아니라 일반적인 Polish metric space 위로 확장되며, 핵심은 인덱스 공간의 기하학적 복잡도와 metric 구조가 된다.

# 2 ULLN

집합들의 모임 $\mathcal{C}$가 유한 집합 $S = \{x_1, \ldots, x_n\}$의 부분집합 $A$를 "pick out"한다는 것은, $A = C \cap S$를 만족하는 어떤 $C \in \mathcal{C}$가 존재한다는 뜻이다. 예를 들어 $\mathbb{R}$ 위에서의 왼쪽 반직선 모임 $\mathcal{C} = \{(-\infty, a] : a \in \mathbb{R}\}$를 생각해보자. 유한 집합 $S = \{1, 3, 5\}$로 놓겠다. $\mathcal{C}$이 어떤 부분집합을 "pick-out" 할 수 있는지 살펴보고자 한다.

- $1 \leq a < 3$인 경우 (예: $a = 2$), $(-\infty, 2] \cap \{1, 3, 5\} = \{1\}$. $\mathcal{C}$는 $C = (-\infty, 1] \in \mathcal{C}$로 $\{1\}$을 pick-out 한다.
- $\{1, 3\}$은 pick out 할 수 없다.

$S$의 부분집합들 중 $\mathcal{C}$로 pick-out 할 수 있는 모임을 다음과 같이 표기하겠다.

$$\mathcal{C} \cap S := \{C \cap S : C \in \mathcal{C}\}$$

위 집합의 크기를 $\Delta^{\mathcal{C}}(S) = |\mathcal{C} \cap S|$라 놓자. 여기서 $\mathcal{C}$을 고정하고 $|S| = n$인 다양한 $S$를 생각해보자. 이는 $\mathcal{C}$가 데이터 $S$를 얼마나 다양하게 조각낼 수 있는지를 보여준다. 집합 클래스 $m^{\mathcal{C}}(n)$는 $\mathcal{C}$의 '조합론적 복잡도(Combinatorial Complexity)'를 측정하는 측도를 정의한다.

$$m^{\mathcal{C}}(n) = \max_{|S|=n} \Delta^{\mathcal{C}}(S)$$

$m^{\mathcal{C}}(n)$ (Shatter Coefficient)는 원소가 $n$개인 불리 해줄 수 있는 최대의 갯수이므로이 값 $2^n$이하다.

$$m^{\mathcal{C}}(n) \leq 2^n$$

**Definition 2.1** (Shattering). $\mathcal{C}$가 $S$를 shatter한다는 것을 $\Delta^{\mathcal{C}}(S) = 2^{|S|}$인 경우로 정의한다.

**Definition 2.2** (VC dimension). 어떤 정수 $m$이 존재하여, 크기가 $m$인 그 어떤 집합도 $\mathcal{C}$가 shatter할 수 없다면 그 정수를 VC dimension of $\mathcal{C}$라 부르고 $V(\mathcal{C})$라고 표기한다.

**Definition 2.3** (VC class). $V(\mathcal{C}) < \infty$인 $\mathcal{C}$를 VC-class라고 한다.

$V(\mathcal{C}) = 5$인 $\mathcal{C}$를 생각하자. 그러면 우리는 $|S| < 5$ 인 집합에 대해서는 $\Delta^{\mathcal{C}}(S) = 2^{|S|}$ 인 $S$을 찾을 수 있다는 뜻이다. 하지만 $|S| = 5$인 어떠한 $S$에 대해서도 $\Delta^{\mathcal{C}}(S) = 2^{|S|}$는 만족시키지 않는다는 뜻이다. 따라서 $V(\mathcal{C})$는 $\mathcal{C}$의 pick-out을 통한 완전분해능력의 상한을 보여준다. 다음 Sauer's Lemma는 VC class $\mathcal{C}$의 분해능의 그 상한이 지수적으로 증가하다가 결국은 다항함수로 증가함을 보여준다.

**Theorem 2.4** (Sauer's Lemma). $V(\mathcal{C}) = k$ 이라면, 모든 $n \geq k$에 대해 다음이 성립한다.

$$m^{\mathcal{C}}(n) \leq \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{k-1}$$

*Sauer's Lemma*는 $m^{\mathcal{C}}(n)$이 $n$에 대해 지수적으로 증가하다가, 특정 지점(*VC-index*)을 넘어서면 다항식(*polynomial*) 수준으로 증가 속도가 급격히 떨어진다는 것을 보여준다.

$|\mathcal{C}| < \infty$ 인 경우 ULLN -> $|\mathcal{C}|$ 이 무한인 경우에 ULLN을 증명하는 방법이 직접적이지 않음
$V(\mathcal{C}) = k$ 인 경우에 ULLN이 어떻게 가능한가?
집합의 개수가 유한하다면, 각 집합 $C$에 대해 성립하는 대수의 법칙(LLN)을 단순히 Union Bound로 묶을 수 있다.

$$P\left(\sup_{C\in\mathcal{C}} |P_n(C) - P(C)| > \epsilon\right) \leq \sum_{C\in\mathcal{C}} P(|P_n(C) - P(C)| > \epsilon)$$

하지만 위 식의 $|\mathcal{C}|$ 자리에 $\infty$가 들어가는 경우 Union Bound를 잡을 수 없다. 개별적으로는 수렴하지만, uniformly 수렴한다고 말할 근거가 없다.
하지만 우리는 경험적 측도의 특성을 이용해 이를 해결하고자 한다. 먼저
이를 위해 Symmetrization을 사용한다. 원래 샘플 $S_n$과 독립적인 ghost sample $S'_n$을 도입하면

$$P\left(\sup_{C} |P_n(C) - P(C)| > \epsilon\right) \leq 2\, P\left(\sup_{C} |P_n(C) - P'_n(C)| > \tfrac{\epsilon}{2}\right)$$

따라서 우변을 bound하는 것으로 충분하다. [증명]
경험적 측도 $P_n$의 정의는

$$P_n(C) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{X_i \in C\}$$

여기서 값 $P_n(C)$를 결정하는 유일한 요소는 집합 $C$가 샘플 점들 $\{X_1, \ldots, X_n\}$ 중 어떤 것들을 포함하는가 즉 pick-out 한 점이 무엇인가하는 것이다.
두 샘플 $S_n$과 $S'_n$을 합친 집합을 $S_{2n} = \{x_1, \ldots, x_{2n}\}$이라 할 때, 어떤 두 집합 $A, B \in \mathcal{C}$가 $S_{2n}$에 대해 동일한 부분집합을 형성한다고 가정해보자

$$A \cap S_{2n} = B \cap S_{2n}$$

이 상태라면, $S_{2n}$에 속한 그 어떤 점 $x_i$에 대해서도 $\mathbb{I}(x_i \in A) = \mathbb{I}(x_i \in B)$가 성립한다.

$\mathcal{C}$ 위에 다음과 같은 동치 관계 $\sim$을 정의한다.

$$A \sim B \iff A \cap S_{2n} = B \cap S_{2n}$$

즉, 두 집합 $A, B \in \mathcal{C}$가 결합된 샘플 $S_{2n}$의 '동일한 점들을 포함할 때' 동치로 본다. 이 관계가 반사성(reflexivity), 대칭성(symmetry), 추이성(transitivity)을 만족함은 자명하므로, $\sim$은 $\mathcal{C}$ 위의 동치 관계다. 이 동치 관계에 의한 몫집합(quotient set)을 $\mathcal{C}/\sim$이라 하면, 각 동치류(equivalence class) $[C]$는

$$[C] = \{A \in \mathcal{C} : A \cap S_{2n} = C \cap S_{2n}\}$$

으로 표현된다.
한편, $\mathcal{C}$가 $S_{2n}$ 위에서 형성할 수 있는 서로 다른 부분집합의 수는 정의에 의해 $m^{\mathcal{C}}(2n)$이하이므로, 동치류의 총 개수 $N$은

$$N = |\mathcal{C}/\sim| \leq m^{\mathcal{C}}(2n)$$

을 만족한다.
이제 각 동치류에서 대표 집합을 하나씩 선택하여 $C_1, C_2, \ldots, C_N$이라 하자. Supremum은 다음과 같이 유한한 Maximum으로 바뀐다.

$$\sup_{C \in \mathcal{C}} |P_n(C) - P'_n(C)| = \max_{j=1,\ldots,N} |P_n(C_j) - P'_n(C_j)|$$

이 "유한한 집합으로의 근사"가 완료되면, 통계학의 가장 강력한 도구인 Union Bound를 쓸 수 있게 된다.

$$P\left(\max_{j \leq m^{\mathcal{C}}(2n)} |P_n(C_j) - P'_n(C_j)| > \epsilon\right) \leq m^{\mathcal{C}}(2n) \cdot \max_j P\left(|P_n(C_j) - P'_n(C_j)| > \epsilon\right)$$

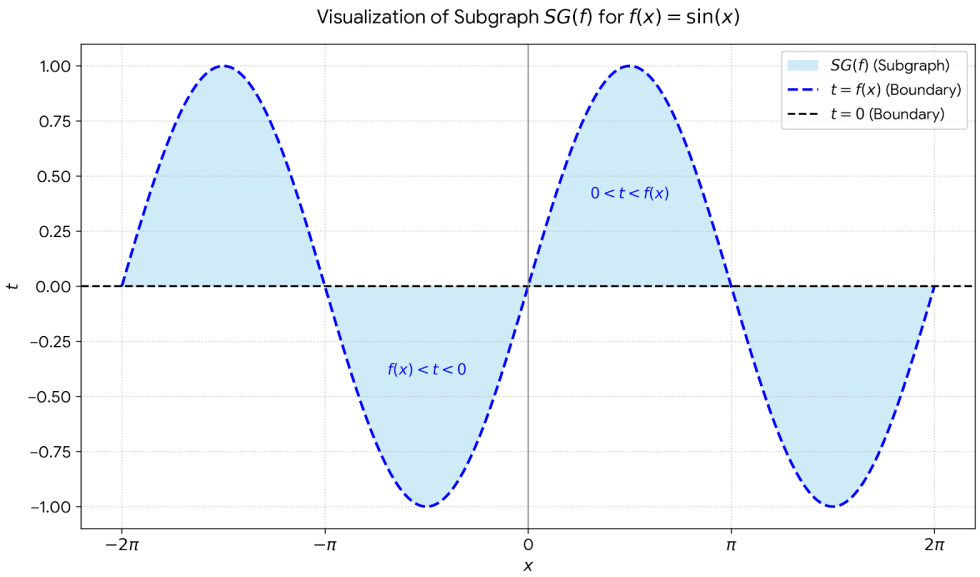
## 2.1 VC function Class

Subgraph (하위 그래프) 정의함수 $f : \mathcal{X} \to \mathbb{R}$의 복잡도를 측정하기 위해, Pollard는 이 함수 아래의 영역을 집합으로 정의한다.

**Definition 2.5.** 함수 $f$의 subgraph는 다음과 같이 정의되는 $\mathcal{X} \times \mathbb{R}$ 공간의 부분집합이다.

$$SG(f) = \{(x, t) \in \mathcal{X} \times \mathbb{R} : 0 < t < f(x) \text{ or } f(x) < t < 0\}$$

**Definition 2.6** (VC-subgraph Class)**.** 함수 클래스 $\mathcal{F}$의 복잡도는 그 함수들의 subgraph들이 이루는 집합 클래스 $\mathcal{C}_{SG} = \{SG(f) : f \in \mathcal{F}\}$의 복잡도로 정의된다. 만약 $\mathcal{C}_{SG}$가 VC-index $k$를 가지는




VC-class라면, 함수 클래스 $\mathcal{F}$를 VC-subgraph class라고 부른다.

**Definition 2.7** (Pick out in Functions). 임의의 이진 라벨(Indicator) 조합 $(b_1, \ldots, b_n) \in \{0,1\}^n$에 대하여, 다음 조건을 만족하는 함수 $f \in \mathcal{F}$가 항상 존재하면, 함수 클래스 $\mathcal{F}$는 점 집합 $P_n$을 'Pick out(또는 Shatter)' 한다고 정의한다.

$$\forall i \in \{1, \ldots, n\}, \quad \mathbb{I}\left((x_i, t_i) \in SG(f)\right) = b_i$$

Pick out in Functions 의 개념은 $n$개의 데이터 포인트 $x_1, \ldots, x_n$과 임의의 값 $t_1, \ldots, t_n$이 주어졌을 때, $\mathcal{F}$의 함수들이 이 $(x_i, t_i)$ 쌍들을 얼마나 자유롭게 분리할 수 있는가를 측정하는 것이다.
$((x_i, t_i), b_i)$ for $i = 1, \cdots, n$ 이 주어져 있다고 하자. 이는 함수 클래스 $\mathcal{F}$가 만족시켜야 할 $n$개의 '제약 조건(Constraints)' 또는 '목표 패턴(Target Pattern)'을 의미한다. $t_i$의 부호에 따라 이 요구사항은 다음과 같이 나누어 설명할 수 있다.

- ($t_i > 0$ 인 경우) Subgraph의 정의상 함수의 값 $f(x_i)$가 반드시 양수이면서 $t_i$보다 커야만 $(x_i, t_i)$가 $SG(f)$ 영역 안에 포함될 수 있다. 즉 $b_i = 1$ 인 경우 $0 < t_i < f(x_i)$ 가 성립해야하고 반대로 $b_i = 0$ 인 경우 $f(x_i) \leq t_i$ 가 성립해야 한다.
- ($t_i < 0$ 인 경우) 임계값 Subgraph의 정의상 함수의 값이 반드시 음수이면서 $t_i$보다 더 작아야 $(x_i, t_i)$가 $SG(f)$ 영역 안에 포함될 수 있다. 즉, $b_i = 1$ 인 경우 $f(x_i) < t_i < 0$ 이 성립해야만 하고, 반대로 $b_i = 0$ 인 경우 $f(x_i) \geq t_i$ 가 성립해야 한다.

이를 통해 보면 $b_i = 1$ 의 의미는 $(x_i, t_i)$ 에서 sign$(f(x_i)) =$ sign$(t_i)$ 이면서 $|t_i| \leq |f(x_i)|$를 만족하는 $f$의 존재성을 의미한다. 종합하면 Pick out in Functions 은 $\mathcal{F}$가 주어진 데이터 $(x_i, t_i)$ for $i = 1, \cdots, n$를 크고 작음을 이용해서 분리할 수 있는지를 나타내는 것이다.
앞서 집합을 이용한 데이터 분리 문제에서 Symmetrization lemma 에 의해 UNNL을 보일 때 $2n$개의 샘플 $S_{2n} = \{x_1, \ldots, x_{2n}\}$에 대한 복잡도를 고려하였다. 이 때 $\mathcal{C}$가 표현할 수 있는 최대 복잡도는 $m^{\mathcal{C}}(2n)$였다. 이와 유사하게 함수 클래스 $\mathcal{F}$에서는, 각 샘플 $x_i$에 대응하는 임의의 높이 값 $t_i$들을 배치한 샘플 $S'_{2n} = (x_1, t_1), \ldots, (x_{2n}, t_{2n})$을 생각한다. 만약 $\mathcal{F}$가 VC-subgraph class (index $k$)라면:어떤 $S'_{2n}$을 가져와도 $\mathcal{F}$가 이 점들을 '함수 아래 영역'으로 분리시켜 만들 수 있는 패턴의 가짓수는 여전히 다항식 $O(n^k)$으로 제한됨을 알 수 있다.

## 2.2 Entropy

**Definition 2.8** (함수공간). $(\mathcal{X}, \mathcal{A})$를 가측 공간(Measurable space)이라 하고, $Q$를 이 공간 위의 측도라 하자.$1 \leq p < \infty$에 대하여, 측도 $Q$에 대한 $L_p$ 공간 $L_p(Q)$는 다음과 같이 정의된다.

$$L_p(Q) = \left\{ g : \mathcal{X} \to \mathbb{R} \;\middle|\; \int |g|^p dQ < \infty \right\}.$$

이때 $g \in L_p(Q)$인 임의의 함수에 대하여, $L_p(Q)$-노름의 $p$제곱은 다음과 같이 작성한다.

$$\|g\|_{p,Q}^p = \int |g|^p dQ$$

$L_p(Q)$ 공간에서 두 함수 $g, h$ 사이의 거리는 두 함수의 차이의 크기인 $d(g, h) = |g - h|_{p,Q}$로 정의하며, $(L_p(Q), d)$는 메트릭 공간임을 알 수 있다.

**Definition 2.9** ($\delta$-Covering). $\delta > 0$에 대하여, 함수들의 모임 $g_1, \ldots, g_N$을 생각하자. 만약 $\mathcal{G}$에 속하는 임의의 함수 $g$를 가져오더라도, $\{g_1, \ldots, g_N\}$ 중에 거리(노름)가 $\delta$ 이하인 원소 $g_j$가 적어도 하나 존재한다면 즉

$$\|g - g_j\|_{p,Q} \leq \delta$$

을 만족하면 그 함수의 모임을 함수 클래스 $\mathcal{G}$의 $\delta$-covering이라고 한다.

**Definition 2.10** (Convering Number & Entropy). $\delta$-covering을 만족하는 반지름이 $\delta$인 공(Ball)들의 최소 개수 $N$을 $\delta$-covering number라고 부르며, 다음과 같이 표기한다.

$$N_p(\delta, \mathcal{G}, Q)$$

(만약 그러한 유한한 개수 $N$이 존재하지 않는다면 $N_p(\delta, \mathcal{G}, Q) = \infty$로 취급) $\delta$-Entropy 는

$$H_p(\delta, \mathcal{G}, Q) = \log N_p(\delta, \mathcal{G}, Q)$$

**Definition 2.11.** 모든 $\delta > 0$에 대하여 엔트로피가 유한할 때 ($H_p(\delta, \mathcal{G}, Q) < \infty$), 함수 클래스 $\mathcal{G}$를 totally bounded 하다고 정의한다.

**Definition 2.12** (Bracketing을 이용한 $L_p(Q)$-엔트로피). 아래 두 조건을 만족하는 두 함수 $g_j^L$(하한)과 $g_j^U$(상한)의 쌍 $[g_j^L, g_j^U]$를 Bracket이라 한다.

- $\|g_j^U - g_j^L\|_{p,Q} \leq \delta \quad (\forall j = 1, \ldots, N)$
- $g_j^L \leq g \leq g_j^U$

집합 $\mathcal{G}$를 완전히 덮을 수 있는 가장 작은 괄호들의 개수 $N$을 의미하며, 기호로는 $N_{p,B}(\delta, \mathcal{G}, Q)$라고 쓴다. Bracketing Number는 함수 집합 $\mathcal{G}$를 완전히 덮을 수 있는 가장 작은 괄호들의 개수 $N$을 의미하며, 기호로는 $N_{p,B}(\delta, \mathcal{G}, Q)$라 쓴다. (Bracketing Entropy는

$$H_{p,B}(\delta, \mathcal{G}, Q) = \log N_{p,B}(\delta, \mathcal{G}, Q)$$

**Definition 2.13** (Entropy for the supremum norm). 다음을 만족하는 $\{g_j\}_{j=1}^N$이 존재하도록 하는 가장 작은 $N$의 값을 $N_\infty(\delta, \mathcal{G})$라 하자.

$$\sup_{g \in \mathcal{G}} \min_{j=1,\ldots,N} |g - g_j|_\infty \leq \delta$$

(만약 이러한 성질을 만족하는 유한 집합이 존재하지 않는다면 $N_\infty(\delta, \mathcal{G}) = \infty$로 둔다.)

**Lemma 2.14.** 모든 $1 \leq p < \infty$ 에 대하여

$$H_p(\delta, \mathcal{G}, Q) \leq H_{p,B}(\delta, \mathcal{G}, Q), \quad (\text{모든 } \delta > 0 \text{ 에 대하여})$$

가 성립한다.게다가, 만약 $Q$가 확률 측도(*probability measure*)라면,

$$H_{p,B}(\delta, \mathcal{G}, Q) \leq H_\infty\left(\frac{\delta}{2}, \mathcal{G}\right), \quad (\text{모든 } \delta > 0 \text{ 에 대하여})$$

가 성립한다.

**Lemma 2.15.** $\mathcal{G}$를 증가함수(*increasing functions*) $g : \mathbb{R} \mapsto [0,1]$들의 클래스라 하고, 유한 집합 $\mathcal{X} \subset \mathbb{R}$의 원소의 개수(*cardinality*)가 $n$이라고 하자. 그러면 모든 $\delta > 0$에 대하여 다음이 성립한다.

$$H_\infty(\delta, \mathcal{G}) \leq \frac{1}{\delta} \log\left(n + \frac{1}{\delta}\right)$$

(proof) 집합 $\mathcal{X}$가 $x_1 \leq \cdots \leq x_n$인 점들로 구성되어 있다고 하자. $\mathcal{G}$에 속하는 각 함수 $g$에 대하여 다음과 같이 정의한다.

$$M_i = \left\lfloor \frac{g(x_i)}{\delta} \right\rfloor, \quad i = 1, \ldots, n,$$

그리고 다음과 같이 둔다.

$$\tilde{g}(x_i) = \delta M_i, \quad i = 1, \ldots, n.$$

그러면 다음이 성립한다.

$$|\tilde{g}(x_i) - g(x_i)| \leq \delta, \quad i = 1, \ldots, n.$$

우리는 $0 \leq M_1 \leq \cdots \leq M_n \leq \lfloor 1/\delta \rfloor$ 이며 $M_i \in \mathbb{N}$ $(i = 1, \ldots, n)$ 임을 알 수 있다. 따라서 임의의 $g$에 대한 가능한 함수 $\tilde{g}$의 총 개수는 다음과 같다[2].

$$\binom{n + \lfloor 1/\delta \rfloor}{\lfloor 1/\delta \rfloor}$$

$\binom{n}{k} \leq \left(\frac{e \cdot n}{k}\right)^k$임을 이용하면

$$\begin{aligned} H_\infty(\delta, \mathcal{G}) &= \log\left(\binom{n + \lfloor 1/\delta \rfloor}{\lfloor 1/\delta \rfloor}\right) \leq \lfloor 1/\delta \rfloor \left(\log(n + \lfloor 1/\delta \rfloor) + 1 - \log(\lfloor 1/\delta \rfloor)\right) \\ &\leq \lfloor 1/\delta \rfloor \log(n + \lfloor 1/\delta \rfloor) \leq \frac{1}{\delta} \log\left(n + \frac{1}{\delta}\right) \end{aligned}$$

**Lemma 2.16.** $\mathcal{G} = \{g : [0,1] \to [0,1],\ |g'| \leq 1\}$ 라 하자. 그러면 어떤 상수 $A$에 대하여 다음이 성립한다.

$$H_\infty(\delta, \mathcal{G}) \leq A\frac{1}{\delta}, \quad \textit{for all } \delta > 0$$

**Lemma 2.17.** 주어진 함수 클래스가 다음과 같을 때,

$$\mathcal{G} = \left\{ g : [0,1] \to [0,1] : \int \left(g^{(m)}(x)\right)^2 dx \leq 1 \right\},$$

[2] $\lfloor 1/\delta \rfloor + 1$ 개 중에서 순서를 고려하지 않고 중복을 허용하여 $n$개를 뽑기로 ${}_nH_r = {}_{n+r-1}C_r$을 이용한다

모든 $\delta > 0$에 대하여 다음이 성립한다.

$$H_\infty(\delta, \mathcal{G}) \leq A\delta^{-1/m}$$

**Proposition 2.18.** 보조정리 *2.5* (*Lemma 2.5*)유클리드 거리(*Euclidean metric*)가 주어진 $\mathbb{R}^d$ 공간에서, 반지름이 $R$인 공 $B_d(R)$은 반지름이 $\delta$인 공들로 최대 다음 개수만큼 있으면 모두 덮을(*Covered*) 수 있다.

$$\left(\frac{4R+\delta}{\delta}\right)^d$$

(*proof*) $\{c_j\}_{j=1}^N \subset B_d(R)$을 서로 다른 두 점 $c_{j_1}$과 $c_{j_2}$ 사이의 거리가 적어도 $\delta$ 이상 떨어져 있도록 하는 가장 큰 집합(*Largest set*, 즉 *Maximal $\delta$-packing set*)이라고 하자.만약 이 조건이 만족된다면, 중심이 $c_j$이고 반지름이 $\delta$인 공들 ($j = 1, \ldots, N$)은 반드시 전체 공 $B_d(R)$을 덮어야 한다. (만약 덮이지 않는 영역이 존재한다면, 그 영역에 거리가 $\delta$ 이상인 새로운 점을 추가할 수 있으므로 '가장 큰 집합'이라는 가정에 모순이 생기기 때문이다.)이제 중심이 $c_j$이고 반지름이 $\delta/4$인 공들을 $B_j$라고 하자 ($j = 1, \ldots, N$). 이 공들의 반지름은 최소 이격 거리인 $\delta$의 절반보다 작기 때문에, $B_1, \ldots, B_N$은 서로소(*Disjoint*, 겹치지 않음)이다. 또한 이들의 합집합은 다음과 같이 약간 확장된 공의 부분집합이 된다.

$$\bigcup_{j=1}^N B_j \subset B_d(R + \delta/4)$$

반지름이 $\rho$인 $d$차원 공의 부피는 $C_d\rho^d$로 표현할 수 있다. 서로소인 공들의 부피의 총합은 전체를 포함하는 큰 공의 부피보다 클 수 없으므로, 다음 부등식이 성립해야 한다.

$$N \cdot C_d(\delta/4)^d \leq C_d(R + \delta/4)^d$$

양변에서 $C_d$를 약분하고 정리하면 다음을 얻는다.

$$N \leq \left(\frac{R+\delta/4}{\delta/4}\right)^d = \left(\frac{4R+\delta}{\delta}\right)^d$$

손실 함수(*Loss function*) 클래스 $\mathcal{F} = \{f_\theta : \theta \in \Theta \subset \mathbb{R}^d\}$ 가 매개변수 $\theta$로 지칭된다고 가정하자. $\Theta$가 반지름 $R$인 유클리드 공 안에 있다면, 이 보조정리를 사용하여 $\Theta$ 위에 총 $N$개의 점으로 이루어진 유한한 $\delta$-*net* $\theta_1, \ldots, \theta_N$을 구성할 수 있다.
만약 손실 함수가 매개변수 $\theta$에 대해 립시츠 연속성($|f_{\theta_1}(x) - f_{\theta_2}(x)| \leq L\|\theta_1 - \theta_2\|$)을 만족한다면,

매개변수 공간의 *Covering*를 이용하여 함수 공간의 *covering*을 만들 수 있다.

$$\|\theta - \theta_j\| \le \delta \implies |f_\theta(x) - f_{\theta_j}(x)| \le L\delta$$

따라서 함수 클래스 $\mathcal{F}$의 $L_\infty$ 엔트로피 상한은 *proposition 2.18*로 부터

$$H_\infty(L\delta, \mathcal{F}) \le d \log\left(\frac{4R+\delta}{\delta}\right)$$

**Theorem 2.19.** $L_2(Q)$ 공간에 속하는 $d$개의 함수 $\psi_1, \dots, \psi_d$를 택하고, 이들의 선형 결합(*Linear combination*)으로 이루어진 함수 클래스 $\mathcal{G}$를 다음과 같이 정의하자.

$$\mathcal{G} = \left\{ g = \sum_{k=1}^{d} \theta_k \psi_k : \theta = (\theta_1, \dots, \theta_d)^T \in \mathbb{R}^d,\ \|g\|_Q \le R \right\}$$

그러면 모든 $\delta > 0$에 대하여 다음이 성립한다.

$$H(\delta, \mathcal{G}, Q) \le d \log\left(\frac{4R+\delta}{\delta}\right)$$

# 3 ULLN

**Theorem 3.1.** 함수 클래스 $\mathcal{G}$와 확률 측도 $P$에 대하여, 다음 조건이 만족된다고 가정하자.

$$H_{1,B}(\delta, \mathcal{G}, P) < \infty, \quad \textit{for all } \delta > 0$$

그러면 함수 클래스 $\mathcal{G}$는 균등 대수의 법칙(*ULLN, Uniform Law of Large Numbers*)을 만족한다. 즉, 경험적 측도 $P_n$에 대해 다음이 성립한다.

$$\sup_{g \in \mathcal{G}} |P_n g - Pg| \xrightarrow{p} 0 \quad \textit{as } n \to \infty$$

(*proof*) 가정에 의해 $H_{1,B}(\epsilon, \mathcal{G}, P) < \infty$이므로, $L_1(P)$ 거리가 $\epsilon$ 이하인 유한개의 브래킷 집합 $\{[u_j, v_j]\}_{j=1}^N$이 존재하여 $\mathcal{G}$ 전체를 덮는다. 즉, 다음이 성립한다.

$$N = \exp\left(H_{1,B}(\epsilon, \mathcal{G}, P)\right) < \infty$$

$$\forall j \in \{1, \dots, N\}, \quad P(v_j - u_j) = \int (v_j(x) - u_j(x)) dP(x) \le \epsilon$$

따라서 임의의 함수 $g \in \mathcal{G}$를 선택하면, 이 함수를 품고 있는 특정 브래킷 인덱스 $j \in 1, \ldots, N$가 반드시 존재하여 다음 부등식을 만족한다.

$$u_j(x) \leq g(x) \leq v_j(x), \quad \forall x$$

$$P_n u_j \leq P_n g \leq P_n v_j \quad \textit{and} \quad P u_j \leq P g \leq P v_j$$

이 성립하므로 오차 항 $P_n g - P g$의 상한(*Upper bound*)을 구하면 다음과 같다.

$$P_n g - P g \leq P_n v_j - P u_j = (P_n v_j - P v_j) + P(v_j - u_j)$$

$P(v_j - u_j) \leq \epsilon$이므로,

$$P_n g - P g \leq (P_n v_j - P v_j) + \epsilon \leq \max_{k=1,\ldots,N} (P_n v_k - P v_k) + \epsilon$$

동일한 논리로 하한(*Lower bound*)을 구하면 다음과 같다.

$$P_n g - P g \geq P_n u_j - P v_j = (P_n u_j - P u_j) - P(v_j - u_j) \geq (P_n u_j - P u_j) - \epsilon$$

$$P_n g - P g \geq \min_{k=1,\ldots,N} (P_n u_k - P u_k) - \epsilon$$

위의 상한과 하한은 임의의 $g \in \mathcal{G}$에 대해 인덱스 $j$에 상관없이 공통적으로 성립하므로, 전체 클래스에 대한 상한(*Supremum*) 기호를 씌워도 부등호가 유지된다.

$$\sup_{g \in \mathcal{G}} |P_n g - P g| \leq \max_{k=1,\ldots,N} |P_n v_k - P v_k| + \max_{k=1,\ldots,N} |P_n u_k - P u_k| + \epsilon$$

이제 $n \to \infty$인 극한을 취하자.여기서 $u_k$와 $v_k$는 $\mathcal{G}$의 원소는 아닐 수 있지만 단일한 가적분 함수(*Integrable functions*)들이며, 브래킷의 총 개수 $N$은 고정된 유한한(*Finite*) 수이다.따라서 통상적인 대수의 법칙(*WLLN*)에 의해, 각각의 $k$에 대하여 다음이 성립한다.

$$\forall k, \quad |P_n v_k - P v_k| \xrightarrow{p} 0 \quad \textit{and} \quad |P_n u_k - P u_k| \xrightarrow{p} 0$$

유한개의 확률 수렴 항들의 최대값(*Maximum*) 역시 *0*으로 확률 수렴하므로,

$$\max_{k=1,\ldots,N} |P_n v_k - P v_k| \xrightarrow{p} 0 \quad \textit{and} \quad \max_{k=1,\ldots,N} |P_n u_k - P u_k| \xrightarrow{p} 0$$

따라서 임의의 $\eta > 0$에 대하여 다음 꼬리 확률의 극한을 상한선으로 제한할 수 있다.

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{g\in\mathcal{G}} |P_n g - Pg| > 2\epsilon\right) = 0$$

**Lemma 3.2** (Symmetrization lemma)**.** 확률 측도 $P$로부터 *i.i.d.* 확률변수 샘플 $X_1, \ldots, X_n$에 대응하는 경험적 측도를 $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$라 하자. 또한, $P_n$과 독립이면서 동일한 분포를 따르는 가상의 경험적 측도(*Ghost empirical measure*)를 $P'_n = \frac{1}{n}\sum_{i=1}^n \delta_{X'_i}$라 하자.함수 클래스 $\mathcal{G}$에 속하는 임의의 함수 $g$와 고정된 상수 $\delta > 0$에 대하여 다음 하한 조건이 만족된다고 가정하자.

$$\mathbf{P}\left(\left|\int g\, d(P_n - P)\right| > \frac{\delta}{2}\right) \leq \frac{1}{2}$$

그러면 다음 부등식이 성립한다.

$$\mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\int g\, d(P_n - P)\right| > \delta\right) \leq 2\,\mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\int g\, d(P_n - P'_n)\right| > \frac{\delta}{2}\right)$$

(*proof*) $\mathbf{X} = (X_1, \ldots, X_n)$이라 하자. 임의의 함수 $g \in \mathcal{G}$에 대하여 사건 $A_g$를 다음과 같이 정의한다.

$$A_g = \left\{\mathbf{X} : \left|\int g\, d(P_n - P)\right| > \delta\right\}$$

또한 전체 클래스에 대한 상한 사건을 $A = \bigcup_{g\in\mathcal{G}} A_g$라 정의한다. 사건 $A$가 발생하였을 때, 오차 조건을 만족하는 $\mathcal{G}$ 내의 어떤 $g^* = g^*(\mathbf{X}) \in \mathcal{G}$가 존재하여 다음을 만족한다.

$$\mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\int g\, d(P_n - P)\right| > \delta\right) \leq \mathbf{P}\left(\left|\int g^* d(P_n - P)\right| > \delta\right) = \mathbf{P}(A_{g^*})$$

그리고 아래 사건을 생각해보자

$$\left\{A_{g^*}\ \textit{and}\ \left|\int g^*\, d(P'_n - P)\right| \leq \frac{\delta}{2}\right\} \in \mathcal{F}(X_1, \cdots, X_n, X'_1, \cdots X'_n)$$

$\mathbf{X}$와 $\mathbf{X}'$이 독립이므로 위 사건의 확률은

$$\begin{aligned}\mathbf{P}\left(A_{g^*}\ \textit{and}\ \left|\int g^*\, d(P'_n - P)\right| \leq \frac{\delta}{2}\right) &= \mathbf{P}_{\mathbf{X}'}\left(\left|\int g^*\, d(P'_n - P)\right| \leq \frac{\delta}{2}\right)\mathbf{P}_{\mathbf{X}}(A_{g^*}) \\ &\geq \frac{1}{2}\mathbf{P}_{\mathbf{X}}\left(\left|\int g^*\, d(P_n - P)\right| \leq \frac{\delta}{2}\right)\end{aligned}$$

삼각부등식에 의해

$$\left|\int g^* \, d(P_n - P)\right| - \left|\int g^* \, d(P'_n - P)\right| \le \left|\int g^* \, d(P_n - P'_n)\right|$$

이므로

$$\left\{\left|\int g^* \, d(P_n - P)\right| > \delta \text{ and } \left|\int g^* \, d(P'_n - P)\right| \le \frac{\delta}{2}\right\} \subset \left\{\left|\int g^* \, d(P_n - P'_n)\right| > \frac{\delta}{2}\right\}$$

$$\begin{aligned}
\mathbf{P}_{\mathbf{X}}\left(\sup_{g\in\mathcal{G}}\left|\int g \, d(P_n - P)\right| > \delta\right) &\le \mathbf{P}_{\mathbf{X}}\left(\left|\int g^* \, d(P_n - P)\right| > \delta\right) \\
&\le 2\,\mathbf{P}_{\mathbf{X},\mathbf{X}'}\left(\left|\int g^* \, d(P_n - P)\right| > \delta \text{ and } \left|\int g^* \, d(P'_n - P)\right| \le \frac{\delta}{2}\right) \\
&\le 2\,\mathbf{P}_{\mathbf{X},\mathbf{X}'}\left(\left|\int g^* \, d(P_n - P'_n)\right| > \frac{\delta}{2}\right) \\
&\le 2\,\mathbf{P}_{\mathbf{X},\mathbf{X}'}\left(\sup_{g\in\mathcal{G}}\left|\int g \, d(P_n - P'_n)\right| > \frac{\delta}{2}\right) \quad (\because g^* \in \mathcal{G})
\end{aligned}$$

**Lemma 3.3** (Symmetrization with Rademacher Sequence). 확률 측도 $P$로부터의 *i.i.d.* 샘플 $\mathbf{X} = (X_1, \ldots, X_n)$에 대응하는 경험적 측도를 $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$라 하자. 또한, $\mathbf{X}$와 독립이면서 각 원소가 다음을 만족하는 *Rademacher Sequence* $\mathbf{W} = (W_1, \ldots, W_n)$을 도입하자.

$$\mathbf{P}(W_i = 1) = \mathbf{P}(W_i = -1) = \frac{1}{2}, \quad i = 1, \ldots, n$$

$\sup_{g\in\mathcal{G}} \|g\| \le R$이고 샘플 크기가 $n \ge 8R^2/\delta^2$를 만족할 때, 임의의 $\delta > 0$에 대하여 다음 부등식이 성립한다.

$$\mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\int g \, d(P_n - P)\right| > \delta\right) \le 4\,\mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^n W_i g(X_i)\right| > \frac{\delta}{4}\right)$$

(*proof*) 임의의 고정된 $g \in \mathcal{G}$에 대하여, $\mathbf{E}[g(X_i)] = \int g \, dP$이므로 경험적 평균의 기댓값은 실제 기댓값과 일치한다. 또한 가정에 의해 $\sup_{g\in\mathcal{G}} \|g\| \le R$이므로 각 확률변수의 분산은 다음과 같이 유계된다.

$$Var(g(X_i)) \le \mathbf{E}[g(X_i)^2] \le \|g\|^2 \le R^2$$

*Chebyshev's inequality*) 적용하면 다음이 성립한다.

$$\mathbf{P}\left(\left|\int g\, d(P_n - P)\right| > \frac{\delta}{2}\right) \leq \frac{Var(\int g\, dP_n)}{(\delta/2)^2} = \frac{\frac{1}{n}Var(g(X_1))}{\delta^2/4} \leq \frac{4R^2}{n\delta^2}$$

여기에 본 정리에 주어진 샘플 크기 조건 $n \geq \frac{8R^2}{\delta^2}$을 대입하면 다음 하한 조건을 만족하게 된다.

$$\frac{4R^2}{n\delta^2} \leq \frac{4R^2}{\left(\frac{8R^2}{\delta^2}\right)\delta^2} = \frac{1}{2}$$

따라서 *Symmetrization lemma* 에 의해

$$\mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\int g\, d(P_n - P)\right| > \delta\right) \leq 2\,\mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n}(g(X_i) - g(X_i'))\right| > \frac{\delta}{2}\right)$$

이때, 각각의 $i$에 대하여 $X_i$와 $X_i'$는 *i.i.d.*이므로 확률변수 $g(X_i) - g(X_i')$의 분포는 $0$을 중심으로 대칭이다. 따라서 랜덤열 $\omega_i \in \{-1, 1\}$들의 벡터 $\omega = (\omega_1, \ldots, \omega_n)$에 대하여

$$(g(X_1) - g(X_1'), \ldots, g(X_n) - g(X_n')) \overset{d}{=} (\omega_1(g(X_1) - g(X_1')), \ldots, \omega_n(g(X_n) - g(X_n')))$$

$$\mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n}(g(X_i) - g(X_i'))\right| > \frac{\delta}{2}\right) = \mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n}W_i(g(X_i) - g(X_i'))\right| > \frac{\delta}{2}\right)$$

한편

$$\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n}W_i(g(X_i) - g(X_i'))\right| \leq \sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n}W_i g(X_i)\right| + \sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n}W_i g(X_i')\right|$$

이므로

$$\begin{aligned}&\mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n}W_i g(X_i)\right| + \sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n}W_i g(X_i')\right| > \frac{\delta}{2}\right)\\ &\leq \mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n}W_i g(X_i)\right| > \frac{\delta}{4}\right) + \mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n}W_i g(X_i')\right| > \frac{\delta}{4}\right)\end{aligned}$$

$$\begin{aligned}\mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\int g\,d(P_n-P)\right|>\delta\right) &\le 2\mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^n W_i(g(X_i)-g(X_i'))\right|>\frac{\delta}{2}\right)\\ &\le 4\,\mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^n W_i g(X_i)\right|>\frac{\delta}{4}\right).\end{aligned}$$

**Lemma 3.4.** *Suppose that for some constants $C_1$ and $C_2$, we have that for each $\gamma\in\mathbf{R}^n$ and all $a>0$,*

$$\mathbf{P}\left(\left|\sum_{i=1}^n W_i\gamma_i\right|\ge a\right)\le C_1\exp\left[-\frac{a^2}{C_2^2\sum_{i=1}^n\gamma_i^2}\right]. \tag{1}$$

*Assume $\sup_{g\in\mathcal{G}}\|g\|_{Q_n}\le R$. There exists a constant $C$ depending only on $C_1$ and $C_2$ such that for all $0\le\epsilon<\delta$, and $K>0$ satisfying*

$$\sqrt{n}(\delta-\epsilon)\ge C\left(\int_{\epsilon/4K}^R H^{1/2}(u,\mathcal{G},Q_n)\,du\vee R\right), \tag{2}$$

*we have*

$$\mathbf{P}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^n W_i g(\xi_i)\right|\ge\delta\wedge\frac{1}{n}\sum_{i=1}^n W_i^2\le K^2\right)\le C\exp\left[-\frac{n(\delta-\epsilon)^2}{C^2R^2}\right]. \tag{3}$$

*Let $\{g_j^s\}_{j=1}^{N_s}$ be a minimal $2^{-s}R$-covering set of $\mathcal{G}$, $s=0,1,\ldots,S$, so that $N_s=N(2^{-s}R,\mathcal{G},Q_n)$. On the set*

$$\left\{\frac{1}{n}\sum_{i=1}^n W_i^2\le K^2\right\},$$

*one can apply the Cauchy–Schwarz inequality to obtain that for*

$$S=\min\{s\ge1:2^{-s}R\le\varepsilon/K\},$$

$$\left|\frac{1}{n}\sum_{i=1}^n W_i\big(g_\theta(\xi_i)-g_\theta^S(\xi_i)\big)\right|\le K\|g_\theta-g_\theta^S\|_{Q_n}\le\varepsilon.$$

*This shows that it suffices to prove the exponential inequality for*

$$P\left(\max_{j=1,\ldots,N_S}\left|\frac{1}{n}\sum_{i=1}^n W_i g_j^S(\xi_i)\right|\ge\delta-\varepsilon\right).$$

*Now use chaining. Write*

$$g_\theta^S = \sum_{s=1}^{S} \bigl(g_\theta^s - g_\theta^{s-1}\bigr).$$

*Note that by the triangle inequality,*

$$\|g_\theta^s - g_\theta^{s-1}\|_{Q_n} \le \|g_\theta^s - g_\theta\|_{Q_n} + \|g_\theta - g_\theta^{s-1}\|_{Q_n} \le 2^{-s}R + 2^{-s+1}R = 3(2^{-s}R).$$

*Let $\eta_s$ be positive numbers satisfying*

$$\sum_{s=1}^{S} \eta_s \le 1.$$

*Then*

$$\begin{aligned}
&P\left(\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n} W_i\bigl(g_\theta^s(\xi_i) - g_\theta^{s-1}(\xi_i)\bigr)\right| \ge \delta - \varepsilon\right)\\
&\le \sum_{s=1}^{S} P\left(\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n} W_i\bigl(g_\theta^s(\xi_i) - g_\theta^{s-1}(\xi_i)\bigr)\right| \ge (\delta-\varepsilon)\eta_s\right)\\
&\le \sum_{s=1}^{S} C_1 \exp\left[2H(2^{-s}R, \mathcal{G}, Q_n) - \frac{n(\delta-\varepsilon)^2\eta_s^2}{9C_2^2\, 2^{-2s}R^2}\right].
\end{aligned}$$

*What is a good choice for $\eta_s$? We take*

$$\eta_s = \frac{6C_2\, 2^{-s}RH^{1/2}(2^{-s}R, \mathcal{G}, Q_n)}{\sqrt{n}\,(\delta-\varepsilon)} \;\vee\; \frac{2^{-s}\sqrt{s}}{8}.$$

*Then indeed, by (3.6),*

$$\begin{aligned}
\sum_{s=1}^{S}\eta_s &\le \sum_{s=1}^{S} \frac{6C_2\, 2^{-s}RH^{1/2}(2^{-s}R, \mathcal{G}, Q_n)}{\sqrt{n}\,(\delta-\varepsilon)} + \sum_{s=1}^{S}\frac{2^{-s}\sqrt{s}}{8}\\
&\le \frac{1}{2} + \frac{1}{2} = 1.
\end{aligned}$$

*Here, we used the bound*

$$\begin{aligned}\sum_{s=1}^{\infty} 2^{-s}\sqrt{s} &\leq 1 + \int_1^{\infty} 2^{-x}\sqrt{x}\,dx \\ &\leq 1 + \int_0^{\infty} 2^{-x}\sqrt{x}\,dx \\ &= 1 + \left(\frac{\pi}{\log 2}\right)^{1/2} \\ &\leq 4.\end{aligned}$$

*Observe that*

$$\eta_s \geq \frac{6C_2\, 2^{-s} R H^{1/2}(2^{-s}R, \mathcal{G}, Q_n)}{\sqrt{n}\,(\delta - \varepsilon)},$$

*so that*

$$H(2^{-s}R, \mathcal{G}, Q_n) \leq \frac{n(\delta-\varepsilon)^2\eta_s^2}{36C_2^2\, 2^{-2s}R^2}.$$

*Thus,*

$$\sum_{s=1}^{S} C_1 \exp\left[2H(2^{-s}R, \mathcal{G}, Q_n) - \frac{n(\delta-\varepsilon)^2\eta_s^2}{9C_2^2\, 2^{-2s}R^2}\right] \leq \sum_{s=1}^{S} C_1 \exp\left[-\frac{n(\delta-\varepsilon)^2\eta_s^2}{18C_2^2\, 2^{-2s}R^2}\right].$$

*Next, invoke that*

$$\eta_s \geq \frac{2^{-s}\sqrt{s}}{8}.$$

*Then*

$$\begin{aligned}\sum_{s=1}^{S} C_1 \exp\left[-\frac{n(\delta-\varepsilon)^2\eta_s^2}{18C_2^2\, 2^{-2s}R^2}\right] &\leq \sum_{s=1}^{S} C_1 \exp\left[-\frac{n(\delta-\varepsilon)^2 s}{1152\, C_2^2 R^2}\right] \\ &\leq \sum_{s=1}^{\infty} C_1 \exp\left[-\frac{n(\delta-\varepsilon)^2 s}{1152\, C_2^2 R^2}\right] \\ &= C_1 \left(1 - \exp\left[-\frac{n(\delta-\varepsilon)^2}{1152\, C_2^2 R^2}\right]\right)^{-1} \exp\left[-\frac{n(\delta-\varepsilon)^2}{1152\, C_2^2 R^2}\right] \\ &\leq 2C_1 \exp\left[-\frac{n(\delta-\varepsilon)^2}{1152\, C_2^2 R^2}\right],\end{aligned}$$

*where in the last inequality, we used the assumption, included in (3.6), that*

$$\frac{n(\delta-\varepsilon)^2}{1152\,C_2^2R^2} \geq \log 2.$$

*Thus, we have shown that*

$$\mathbf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n}W_i g(\xi_i)\right| \geq \delta \ \wedge\ \frac{1}{n}\sum_{i=1}^{n}W_i^2 \leq K^2\right) \leq 2C_1 \exp\left[-\frac{n(\delta-\varepsilon)^2}{1152\,C_2^2R^2}\right]. \tag{3.7}$$

**Theorem 3.5** (Universal Entropy Bound for VC-subgraph Classes)**.** 함수 클래스 $\mathcal{F}$가 *VC-index* $V$를 가지는 *VC-subgraph class*이고, $F$가 이 클래스의 포락 함수(*Envelope function*)라고 하자. 그러면 임의의 확률 측도 $Q$와 $r \geq 1$에 대하여, 피복 수는 다음을 만족한다.

$$N(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)) \leq KV(16e)^V\left(\frac{1}{\epsilon}\right)^{r(V-1)}$$

(단, $0<\epsilon<1$ 이고, $K$는 절대 상수)

# 4 Donsker Class

**Definition 4.1.** 함수공간 $\mathcal{F}$가 $P$-Donsker라는 것은 다음이 성립함을 의미한다.

$$\mathbb{G}_n \Rightarrow \mathbb{G}_P \quad \text{in } \ell^\infty(\mathcal{F})$$

여기서 $\ell^\infty(\mathcal{F}) = H : \mathcal{F} \to \mathbb{R} \mid \sup_{f\in\mathcal{F}}|H(f)| < \infty$이다. 극한 과정 $\mathbb{G}_P$는 $\ell^\infty(\mathcal{F})$ 상의 tight Gaussian process인 Brownian bridge process 로, 평균이 0이고 공분산 함수가 다음과 같다.

$$E[\mathbb{G}_P f \cdot \mathbb{G}_P g] = P(fg) - (Pf)(Pg), \quad \forall f,g \in \mathcal{F}$$

**Theorem 4.2.** $\mathbb{G}_n \Rightarrow \mathbb{G}_P$ *in* $\ell^\infty(\mathcal{F})$가 성립하기 위한 필요충분조건은 다음 두 가지이다.

- 유한차원 수렴 (*Finite-dimensional convergence*): 임의의 유한한 부분집합 $f_1,\ldots,f_m \subset \mathcal{F}$에 대하여, $(\mathbb{G}_n f_1,\ldots,\mathbb{G}_n f_m)$이 $(\mathbb{G}_P f_1,\ldots,\mathbb{G}_P g_m)$으로 다변량 중심극한정리(*CLT*)에 의해 약수렴한다.
- 점근적 균등연속성 (*Asymptotic equicontinuity*): 세미-메트릭 공간 $(\mathcal{F},\rho_P)$ 상에서 $G_n$이 *Asymptotic equicontinuity*다.

**Definition 4.3** (세미-메트릭 $\rho_P$). 함수공간 $\mathcal{F}$ 위의 세미-메트릭 $\rho_P$는 다음과 같이 $L_2(P)$ 노름으로 정의된다.

$$\rho_P(f,g) = \|f-g\|_{P,2} = \left(P(f-g)^2\right)^{1/2} = \left(\int (f-g)^2 dP\right)^{1/2}$$

**Definition 4.4** (Asymptotic equicontinuity). 함수공간 $\mathcal{F}$가 세미-메트릭 $\rho_P$에 대해 점근적 균등연속성을 만족한다는 것은 임의의 $\eta > 0$에 대하여 다음이 성립함을 의미한다.

$$\lim_{\delta \downarrow 0} \limsup_{n\to\infty} P\left(\sup_{\rho_P(f,g)<\delta} |\mathbb{G}_n f - \mathbb{G}_n g| > \eta\right) = 0$$

(엄밀하게는 확률의 외측도(outer probability) $P^*$으로 정의한다.)

**Definition 4.5** ($\varepsilon$-net). 세미-메트릭 공간 $(\mathcal{F}, L_2(P))$가 주어졌을 때, 유한집합 $T = f_1, \ldots, f_N \subset \mathcal{F}$가 $\mathcal{F}$의 $\varepsilon$-net이라는 것은 다음 조건을 만족함을 의미한다.

$$\forall f \in \mathcal{F}, \quad \exists f_j \in T \quad \text{such that} \quad \|f - f_j\|_{P,2} \le \varepsilon$$

**Definition 4.6.** 각 정수 $k \ge 0$에 대하여, $T_k = f_{k,1}, \ldots, f_{k,N_k} \subset \mathcal{F}$를 반경 $\varepsilon_k$에 대응하는 최소 크기의 $\varepsilon_k$-net으로 정의한다. 이때 원소의 개수는 다음을 만족한다.

$$N_k = N(\varepsilon_k, \mathcal{F}, L_2(P))$$

**Definition 4.7** (Projection 사상). 각 레벨 $k$에 대해, 함수공간 $\mathcal{F}$에서 유한집합 $T_k$로의 사영 사상 $\pi_k : \mathcal{F} \to T_k$를 다음과 같이 정의한다.

$$\pi_k(f) = \arg\min_{g \in T_k} \|f - g\|_{P,2}$$

이 사상은 최소 원소의 성질에 의하여 다음 부등식을 필연적으로 만족한다.

$$\|f - \pi_k(f)\|_{P,2} \le \varepsilon_k, \quad \forall f \in \mathcal{F}$$

**Theorem 4.8** (Telescope decomposition). 임의의 고정된 최대 분해 해상도 레벨 $K \in \mathbb{N}$에 대하여, 임의의 함수 $f \in \mathcal{F}$는 다음과 같이 각 스케일별 사영점들의 차이(*increment*)의 합으로 항등 분해될 수 있다.

$$f = \pi_0(f) + \sum_{k=1}^{K} \big(\pi_k(f) - \pi_{k-1}(f)\big) + \big(f - \pi_K(f)\big)$$

경험과정 연산자 $\mathbb{G}_n$은 선형 연산자이므로 위의 대수적 분해에 $\mathbb{G}_n$을 직접 적용하면 다음의 체이닝 분해식(*Chaining decomposition*)을 얻는다.

$$\mathbb{G}_n f = \mathbb{G}_n \pi_0(f) + \sum_{k=1}^{K} \mathbb{G}_n\big(\pi_k(f) - \pi_{k-1}(f)\big) + \mathbb{G}_n\big(f - \pi_K(f)\big)$$

양변에 전범위 상한(*supremum*)을 취하면 삼각부등식에 의해 다음이 성립한다.

$$\sup_{f\in\mathcal{F}} |\mathbb{G}_n f| \le \sup_{f\in\mathcal{F}} |\mathbb{G}_n \pi_0(f)| + \sum_{k=1}^{K} \sup_{f\in\mathcal{F}} \big|\mathbb{G}_n\big(\pi_k(f) - \pi_{k-1}(f)\big)\big| + \sup_{f\in\mathcal{F}} |\mathbb{G}_n\big(f - \pi_K(f)\big)|$$

임의의 초기 반경 $\sigma > 0$에 대하여, 다음과 같이 공비가 $1/2$인 이진 등비수열(*dyadic scale*) $\varepsilon_k$를 구성한다.

$$\varepsilon_k = 2^{-k}\sigma, \quad k = 0, 1, 2, \ldots$$

임의의 $f \in \mathcal{F}$에 대하여, 인접 레벨 사영점들 사이의 거리는 $\varepsilon_{k-1}$ 단위로 제어된다.

$$\|\pi_k(f) - \pi_{k-1}(f)\|_{P,2} \le 3\varepsilon_k = \frac{3}{2}\varepsilon_{k-1}$$

왜냐하면 $L_2(P)$ 노름 상에서 삼각부등식과 사영 사상의 정의식 $\|f - \pi_k(f)\|_{P,2} \le \varepsilon_k$를 적용할 수 있기 때문이다.

$$\begin{aligned} \|\pi_k(f) - \pi_{k-1}(f)\|_{P,2} &\le \|\pi_k(f) - f\|_{P,2} + \|f - \pi_{k-1}(f)\|_{P,2} \\ &\le \varepsilon_k + \varepsilon_{k-1} \\ &= \varepsilon_k + 2\varepsilon_k = 3\varepsilon_k = \frac{3}{2}\varepsilon_{k-1} \end{aligned}$$

한편 증분 함수 쌍의 전체 집합은 유한집합이며, 그 *cardinality*는 다음의 상한을 가진다.

$$\#\left\{\pi_k(f) - \pi_{k-1}(f) : f \in \mathcal{F}\right\} \le N_k N_{k-1}$$

**Theorem 4.9.** (*Sub-Gaussian Maximal Inequality*)평균이 *0*이고 $L_2(P)$ 노름이 $s$ 이하인 가우시안 또는 아-가우시안(*sub-Gaussian*) 확률변수들 $Z_1, \ldots, Z_m$에 대하여 다음 부등식이 성립한다.

$$E \max_{1\le j\le m} |Z_j| \le C \cdot s\sqrt{\log m}$$

여기서 $C > 0$는 범용 상수이다.

**Theorem 4.10.** 각 레벨 $k$에서의 경험과정 증분의 상한 기댓값은 다음과 같이 유도된다.

$$E \sup_{f \in \mathcal{F}} |\mathbb{G}_n(\pi_k(f) - \pi_{k-1}(f))| \lesssim \varepsilon_{k-1} \sqrt{\log N(\varepsilon_k, \mathcal{F}, L_2(P))}$$

각 증분의 $L_2(P)$ 크기는 $s \asymp \varepsilon_{k-1}$이며, 가능한 총 개수는 $m \leq N_k N_{k-1}$이다. 정리 *7.1*을 해당 유한 집합에 적용하면 다음과 같다.

$$E \sup_{f \in \mathcal{F}} |\mathbb{G}_n(\pi_k(f) - \pi_{k-1}(f))| \leq C \cdot \left( \frac{3}{2} \varepsilon_{k-1} \right) \sqrt{\log(N_k N_{k-1})}$$

이 때 *Covering number*는 반경이 작아질수록 단조 증가하므로 $N_{k-1} \leq N_k$가 성립하며, 이에 따라 $\log(N_k N_{k-1}) \leq \log(N_k^2) = 2 \log N_k$가 유도된다. 이를 대입하여 정리하면 최종적으로 아래의 바운드를 얻는다.

$$E \sup_{f \in \mathcal{F}} |\mathbb{G}_n(\pi_k(f) - \pi_{k-1}(f))| \leq 3\sqrt{2} C \cdot \varepsilon_{k-1} \sqrt{\log N_k} \lesssim \varepsilon_{k-1} \sqrt{\log N(\varepsilon_k, \mathcal{F}, L_2(P))} \quad \square$$

선택된 스케일의 특성상 $\varepsilon_k = 2^{-k}\sigma$이므로, 인접 격자의 간격은 다음과 같다.

$$\varepsilon_{k-1} - \varepsilon_k = 2^{-(k-1)}\sigma - 2^{-k}\sigma = 2^{-k}\sigma(2-1) = \varepsilon_k$$

이를 변형하면 $\varepsilon_{k-1} = 2\varepsilon_k = 2(\varepsilon_{k-1} - \varepsilon_k)$이 성립한다. 이 식을 수열의 합에 대입한다.

$$\sum_{k=1}^{K} \varepsilon_{k-1} \sqrt{\log N(\varepsilon_k)} = 2 \sum_{k=1}^{K} \sqrt{\log N(\varepsilon_k)} (\varepsilon_{k-1} - \varepsilon_k)$$

함수 $u \mapsto \sqrt{\log N(u, \mathcal{F}, L_2(P))}$는 $u$에 대해 감소하는 단조함수이므로, 각 하위 구간 $[\varepsilon_k, \varepsilon_{k-1}]$ 내에서의 최댓값은 왼쪽 끝점인 $u = \varepsilon_k$에서 나타난다. 따라서 직사각형 근사에 의해 다음 부등식이 성립한다.

$$\sqrt{\log N(\varepsilon_k)} (\varepsilon_{k-1} - \varepsilon_k) \leq \int_{\varepsilon_k}^{\varepsilon_{k-1}} \sqrt{\log N(u, \mathcal{F}, L_2(P))}\, du$$

해당 등식을 $k = 1$부터 $K$까지 연쇄적으로 더하면 적분 구간이 계산된다.

$$\sum_{k=1}^{K} \sqrt{\log N(\varepsilon_k)} (\varepsilon_{k-1} - \varepsilon_k) \leq \sum_{k=1}^{K} \int_{\varepsilon_k}^{\varepsilon_{k-1}} \sqrt{\log N(u)}\, du = \int_{\varepsilon_K}^{\varepsilon_0} \sqrt{\log N(u)}\, du$$

양변에 앞서 추출한 계수 *2*를 곱하고 $\varepsilon_0 = \sigma$, $\varepsilon_K = 2^{-K}\sigma$를 적용하면 원하는 엔트로피 적분 상한이

유도된다.

$$\sum_{k=1}^{K} \varepsilon_{k-1} \sqrt{\log N(\varepsilon_k, \mathcal{F}, L_2(P))} \leq 2 \int_{2^{-K}\sigma}^{\sigma} \sqrt{\log N(u, \mathcal{F}, L_2(P))}\, du$$

이진 스케일의 증분 복잡도의 합은 다음과 같이 리만 합(*Riemann sum*) 구조를 통해 연속 엔트로피 적분으로 상한이 설정된다.

$$\sum_{k=1}^{K} \varepsilon_{k-1} \sqrt{\log N(\varepsilon_k, \mathcal{F}, L_2(P))} \leq 4 \int_{2^{-K}\sigma}^{\sigma} \sqrt{\log N(u, \mathcal{F}, L_2(P))}\, du$$

**Theorem 4.11.** *envelope function* $F = \sup_{f \in \mathcal{F}} |f|$가 적분가능일때 다음 부등식이 성립한다.

$$E^* \sup_{f \in \mathcal{F}} |\mathbb{G}_n f| \leq K \int_0^{\sigma} \sqrt{\log N(u, \mathcal{F}, L_2(P))}\, du + \Delta_n(K)$$

$\Delta_n(K)$는 정밀도 해상도 $K$를 충분히 크게 잡고 $n \to \infty$로 보낼 때 오차 항이 $o(1)$로 소멸함을 나타내는 나머지 항(*remainder term*)이다.

$$E^* \sup_{f \in \mathcal{F}} |\mathbb{G}_n f| \leq E^* \max_{g \in T_0} |\mathbb{G}_n g| + 2C \int_{2^{-K}\sigma}^{\sigma} \sqrt{\log N(u)}\, du + E^* \sup_{f \in \mathcal{F}} |\mathbb{G}_n (f - \pi_K f)|$$

여기서 $\pi_K f$는 $\|f - \pi_K f\|_{P,2} \leq 2^{-K}\sigma$이다. $K$를 고정시킨 상태에서 $n \to \infty$ 일 때, *remainder term*은 대수의 법칙 및 중심극한정리의 확장에 의해 다음과 같이 무시 가능한 수준으로 제어된다.

$$\lim_{K \to \infty} \limsup_{n \to \infty} E^* \sup_{f \in \mathcal{F}} |\mathbb{G}_n (f - \pi_K f)| = 0$$

따라서 잔여 오차는 $o(1)$로 소멸하며, 순수 엔트로피 적분이 지배 항으로 남는다.

**Definition 4.12.** 임의의 국소 정밀도 반경 $\delta > 0$에 대하여, 오차 함수들의 집합인 Localized Class $\mathcal{F}_\delta$를 다음과 같이 정의한다.

$$\mathcal{F}_\delta = \{f - g : f, g \in \mathcal{F},\ \|f - g\|_{P,2} < \delta\}$$

**Theorem 4.13.** 만약 다음의 엔트로피 적분 조건이 성립한다면,

$$\int_0^1 \sqrt{\log N(u, \mathcal{F}, L_2(P))}\, du < \infty$$

$\mathcal{F}$는 점근적 균등연속성을 만족하며, 최종적으로 $P$*-Donsker class*가 된다.

(*proof*) 국소 공간 $\mathcal{F}_\delta$에 속한 원소들의 $L_2(P)$ 노름의 상한은 정의상 $\delta$이다. 따라서 *Pollard* 체이닝 부등식(정리 *10.1*)의 초기 스케일 반경을 $\sigma = \delta$로 직접 지정하여 대입할 수 있다.

$$E^* \sup_{h \in \mathcal{F}_\delta} |\mathbb{G}_n h| \leq 2C \int_0^\delta \sqrt{\log N(u, \mathcal{F}, L_2(P))}\, du + o(1)$$

조건에 의해 엔트로피 적분 함수가 $0$ 근방에서 가적분(*integrable*)하므로, 적분의 절대연속성에 의하여 다음 극한이 보장된다.

$$\lim_{\delta \downarrow 0} \int_0^\delta \sqrt{\log N(u, \mathcal{F}, L_2(P))}\, du = 0$$

이를 마르코프 부등식(*Markov's inequality*)에 대입하면 임의의 임계치 $\eta > 0$에 대하여 다음이 유도된다.

$$\begin{aligned} \lim_{\delta \downarrow 0} \limsup_{n \to \infty} P^* \left( \sup_{\rho_P(f,g) < \delta} |\mathbb{G}_n f - \mathbb{G}_n g| > \eta \right) &\leq \lim_{\delta \downarrow 0} \limsup_{n \to \infty} \frac{E^* \sup_{h \in \mathcal{F}_\delta} |\mathbb{G}_n h|}{\eta} \\ &\leq \frac{2C}{\eta} \lim_{\delta \downarrow 0} \int_0^\delta \sqrt{\log N(u, \mathcal{F}, L_2(P))}\, du = 0 \end{aligned}$$

# 5 Appendix

**Proposition 5.1.** 임의의 양의 정수 $k \in \mathbb{N}$에 대하여 다음 부등식이 성립한다.

$$k! \geq \left( \frac{k}{e} \right)^k$$

(*Proof*) 단계 *1*: 로그 변환을 통한 동치 관계 정립자연로그 함수 $f(x) = \ln x$는 구간 $(0, \infty)$에서 단조증가(*monotonic increasing*)하므로, 주어진 부등식의 양변에 자연로그를 취해도 부등호의 방향은 보존된다.

$$\ln(k!) \geq \ln \left( \frac{k}{e} \right)^k$$

로그의 성질에 의해 좌변과 우변은 각각 다음과 같이 전개된다.

$$\text{좌변: } \ln(k!) = \ln \left( \prod_{i=1}^k i \right) = \sum_{i=1}^k \ln i$$

$$\text{우변: } \ln \left( \frac{k}{e} \right)^k = k \ln \left( \frac{k}{e} \right) = k \ln k - k \ln e = k \ln k - k$$

따라서 본 문제는 다음 부등식을 증명하는 것과 동치이다.

$$\sum_{i=1}^{k} \ln i \geq k \ln k - k$$

단계 *2*: 정적분을 이용한 하한(*Lower bound*) 유도함수 $f(x) = \ln x$는 구간 $[1, k]$에서 단조 증가하므로, 임의의 양의 정수 $i \in 2, 3, \ldots, k$에 대하여 다음이 성립한다.

$$\forall x \in [i-1, i], \quad \ln x \leq \ln i$$

양변을 구간 $[i-1, i]$에 대해 적분하면 다음과 같다.

$$\int_{i-1}^{i} \ln x \, dx \leq \int_{i-1}^{i} \ln i \, dx = \ln i$$

위 부등식을 $i = 2$부터 $k$까지 각각 더하면(*Summation*) 다음과 같다.

$$\sum_{i=2}^{k} \int_{i-1}^{i} \ln x \, dx \leq \sum_{i=2}^{k} \ln i$$

정적분의 연쇄 성질과 $\ln 1 = 0$임을 이용하면 위 식은 다음과 같이 정리된다.

$$\int_{1}^{k} \ln x \, dx \leq \sum_{i=1}^{k} \ln i$$

단계 *3*: 적분 계산 및 최종 부등식 도출좌변의 정적분을 부분적분법(*Integration by parts*)을 사용하여 계산한다.

$$\begin{aligned} \int_{1}^{k} \ln x \, dx &= \Big[ x \ln x - x \Big]_{1}^{k} \\ &= (k \ln k - k) - (1 \ln 1 - 1) \\ &= k \ln k - k + 1 \end{aligned}$$

단계 *2*의 결과에 대입하면 다음을 얻는다.

$$\sum_{i=1}^{k} \ln i \geq k \ln k - k + 1$$

이때 $1 > 0$이므로 다음이 당연히 성립한다.

$$\sum_{i=1}^{k} \ln i \geq k \ln k - k + 1 > k \ln k - k$$

따라서 로그 변환 전의 동치 관계에 의해 다음 부등식이 성립한다.

$$k! \geq \left(\frac{k}{e}\right)^k$$

**Proposition 5.2.**

$$\binom{n}{k} \leq \left(\frac{e \cdot n}{k}\right)^k$$

(*proof*)

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \leq \frac{n^k}{k!} \leq \frac{n^k}{(k/e)^k} = \left(\frac{e \cdot n}{k}\right)^k$$

두 번째 등식은 *Proposition 3.1* 에 의한 것이다.