

# 통계공부와 관련된 글들

Seoncheol Park

2017-09-07



## 차 례



# 일러두기

John W. Tukey는 이렇게 말했다.

“통계학은 과학이라는 것이 내 견해다. 통계학은 더이상 수학이나 물리학, 화학 또는 경제학의 한 부류가 아니다.”

각 장은 독립된 구성으로 되어 있으며, 한 권 이상의 책들을 참고문헌으로 하여 그들의 정의 및 표현을 따라가는 방식으로 구성되어 있다. 따라서 각 장마다 표현 및 한국어 용어 번역이 상이할 수 있다.



편 I

# Basic Concepts





# 1 기본적인 수학 개념들

이 장에서는 앞으로 다룰 내용을 이해하기 위해 필요한 기본적인 수학 개념을 정리하였다.

## 1.1 집합론 (set theory)

### 1.1.1 카디널리티 (cardinality)

**카디널리티 (cardinality)**는 집합의 원소의 갯수를 세기 위해 도입되었다. 유한집합에서는 원소의 갯수를 세는 것이 어렵지 않지만, 무한집합의 경우는 갯수를 세는 것이 문제가 될 수 있다. 집합  $A$ 가 주어졌을 때, 그것의 카디널리티를  $|A|$ 로 쓰도록 하자. 자연수 집합의 카디널리티는 특별히  $|\mathbb{N}| = \aleph_0$ 로 쓰며 **알레프-널 (aleph-null)**로 부른다.

**Definition 1.1.1** (카디널리티가 같다). 두 집합  $A, B$  사이에 전단사(bijection, 일대일 대응) 관계가 성립할 때, 두 집합의 카디널리티가 같다고 정의하고,  $|A| = |B|$ 로 표기한다.  $\square$

## 1.2 연산자들과 노름 (operators and norms)

### 1.2.1 직합 (direct sum)

**Definition 1.2.1** (직합). 크기  $m \times n$ 인 행렬  $\mathbf{A}$ 와  $p \times q$ 인 행렬  $\mathbf{B}$ 가 있을 때 이들의 직합 (direct sum)은

$$\mathbf{A} \oplus \mathbf{B} = \begin{bmatrix} \mathbf{A} & 0 \\ 0 & \mathbf{B} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & b_{11} & \cdots & b_{1q} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & b_{p1} & \cdots & b_{pq} \end{bmatrix}$$

□

**Example 1.2.1** (직합의 예).

$$\begin{bmatrix} 1 & 3 & 2 \\ 2 & 3 & 1 \end{bmatrix} \oplus \begin{bmatrix} 1 & 6 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 2 & 0 & 0 \\ 2 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

□

### 1.2.2 크로네커 곱 (Kronecker product)

**Definition 1.2.2** (크로네커 곱). 크기  $m \times n$ 인 행렬  $\mathbf{A}$ 와  $p \times q$ 인 행렬  $\mathbf{B}$ 가 있을 때 이들의 크로네커 곱 (Kronecker product)은

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}$$

□

**Example 1.2.2** (크로네커 곱의 예). 다음은 크로네커 곱의 한 예이다.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \otimes \begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix} = \begin{bmatrix} 1 \cdot 0 & 1 \cdot 5 & 2 \cdot 0 & 2 \cdot 5 \\ 1 \cdot 6 & 1 \cdot 7 & 2 \cdot 6 & 2 \cdot 7 \\ 3 \cdot 0 & 3 \cdot 5 & 4 \cdot 0 & 4 \cdot 5 \\ 3 \cdot 6 & 3 \cdot 7 & 4 \cdot 6 & 4 \cdot 7 \end{bmatrix} = \begin{bmatrix} 0 & 5 & 0 & 10 \\ 6 & 7 & 12 & 14 \\ 0 & 5 & 0 & 20 \\ 18 & 21 & 24 & 28 \end{bmatrix}$$

□

### 1.2.3 텐서곱 (tensor product)

### 1.2.4 데카르트 곱 (Cartesian product)

두 개의 집합  $A, B$ 가 있을 때, 이들의 **데카르트 곱 (Cartesian product)**  $A \times B$ 는

$$A \times B = \{(a, b) | a \in A \text{ and } b \in B\}$$

로 정의된다.

### 1.2.5 노름 (norm)

#### 1.2.5.1 벡터 노름 (vector norm)

**Definition 1.2.3** (노름과 노름공간). 벡터공간  $X$ 에서 다음 세 조건들이 만족되면 함수  $\|\cdot\|$ 을 **노름 (norm)**이라 하고 또한 벡터공간  $X$ 를 **노름공간 (normed space)**라 한다.

1. 임의의  $\mathbf{x} \in X$ , where  $\|\mathbf{x}\| \geq 0$ 이며  $\|\mathbf{x}\| = 0$ 이기 위한 필요충분조건은  $\mathbf{x} = \mathbf{0}$ 이다.
2. 임의의  $\mathbf{x}, \mathbf{y} \in X$ 에 대해

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$$

가 성립한다.

3. 임의의 스칼라  $\alpha$ 와 임의의  $\mathbf{x} \in X$ 에 대해

$$\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$$

가 성립한다.

□

### 1.2.5.2 행렬 노름 (matrix norm)

### 1.2.6 거리 (metric)

**Definition 1.2.4** (거리). 집합  $X$ 의 두 원소들  $x, y$ 에 대해 값  $d(x, y)$ 를 할당하는 함수  $d(\cdot, \cdot)$ 가 다음 세 조건들을

1. 임의의  $x, y \in X$ 에 대해서  $d(x, y) \geq 0$ 이다. 여기서  $d(x, y) = 0$ 이기 위한 필요충분 조건은  $x = y$ 이다.
2. 임의의  $x, y \in X$ 에 대해서  $d(x, y) = d(y, x)$ 이다.
3. 임의의  $x, y, z \in X$ 에 대해서  $d(x, z) \leq d(x, y) + d(y, z)$ 가 성립한다.

□

만약  $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ 라 하면 노름공간이 거리공간이 된다. 그러나 모든 거리공간이 노름공간으로부터 유도되는 것은 아니다.

## 1.3 행렬의 분해 (matrix decomposition)

### 1.3.1 고유값 분해 (eigenvalue decomposition)

고유값 분해는 행렬  $A$ 가  $n \times n$  정방행렬일 때만 적용 가능하다.

### 1.3.2 스펙트럼 분해 (spectral decomposition)

$p \times p$  대칭행렬  $A$ 에 대한 **스펙트럼 분해 (spectral decomposition)**는 다음과 같다.  $p \times p$  대칭행렬  $A$ 는 직교행렬  $P$ 에 의해 **대각화 (diagonalization)**된다고 한다.

$$A = P\Lambda P^T = \sum_{i=1}^p \lambda_i e_i e_i^T.$$

이때  $PP^T = P^T P = I$ 를 만족하는 직교행렬  $P$ 는  $P = [e_1, \dots, e_p]$ 로 이루어지며,  $\Lambda$ 는  $A$ 의 고유값 (eigenvalue) 들로만 이루어진 대각행렬 (diagonal matrix)

$$\Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p \end{bmatrix}$$

이다. 대각행렬  $\Lambda$ 는  $P^T A P = \Lambda$ 이다.

### 1.3.3 특이값분해(SVD)

**특이값분해 (singular value decomposition, SVD)**는  $m \times n$  직사각형 행렬  $A$ 에 대해 스펙트럼 분해를 일반화한 것이다.  $A$ 의 특이값 분해는 다음과 같다.

$$A = U \Sigma V^T.$$

이 때

- $U$ :  $A$ 의 left singular vector로 이루어진  $m \times m$  직교행렬 (orthogonal matrix)
- $\Sigma$ : 주 대각성분이  $\sqrt{\lambda_i}$ 로 이루어진  $m \times n$  직사각 대각행렬 (diagonal matrix)
- $V$ :  $A$ 의 right singular vector로 이루어진  $n \times n$  직교행렬 (orthogonal matrix)

행렬  $A$ 의 계수(rank)가  $k$ 라고 할 때,

- $U = [u_1, \dots, u_k, \dots, u_m]$ 는  $AA^T$ 를 고유값분해 (eigenvalue decomposition)로 직교대각화하여 얻은  $m \times m$  직교행렬 (orthogonal matrix)이며, 특히  $[u_1, \dots, u_k]$ 를 **좌특이벡터 (left singular vector)**라고 한다.
- $V = [v_1, \dots, v_k, \dots, v_n]$ 는  $A^T A$ 를 고유값분해로 직교대각화하여 얻은  $n \times n$  직교행렬이며, 특히  $[v_1, v_2, \dots, v_k]$ 를 **우특이벡터 (right singular vector)**라고 한다.
- $\Sigma$ 는  $A^T A$ 의 0이 아닌 고유값이  $\lambda_1, \lambda_2, \dots, \lambda_k$ 일 때  $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_k}$ 를 대각성분으로 가지고 나머지 성분을 0으로 갖는  $m \times n$  직사각 **대각행렬 (diagonal matrix)**이다.

$$\Sigma = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & \cdots & \sqrt{\lambda_k} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

즉  $A$ 를 다시 쓰면

$$A = \begin{bmatrix} u_1 & \cdots & u_k & \cdots & u_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & \cdots & \sqrt{\lambda_k} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ \vdots \\ V_k^T \\ \vdots \\ V_n^T \end{bmatrix}$$

이다. 위 식에서 **특이값 (singular value)**는  $\sigma_i^2 = \lambda_i$ 로부터  $\sigma_i = \sqrt{\lambda_i}$ 가 된다. 참고로  $U, V$ 가 직교행렬이면  $UU^T = I, VV^T = I$ 가 성립한다.

### 1.3.4 특이값분해와 고유값분해의 관계

$m \times n$  행렬  $A$ 의 특이값분해의  $U$ 는  $AA^T$ 의 고유벡터이고,  $V$ 는  $A^T A$ 의 고유벡터이며,  $A$ 의 0이 아닌 특이값들의 제곱  $\Sigma \Sigma^T, \Sigma^T \Sigma$ 는  $AA^T, A^T A$ 의 고유값과 같음을 알 수 있다. 참고로  $\sigma_i = \sqrt{\lambda_i}$ 이므로  $\Sigma \Sigma^T$  또는  $\Sigma^T \Sigma = \lambda_i$ 이다.

$$\begin{aligned}
U &= AA^T \\
&= (U\Sigma V^T)(U\Sigma V^T)^T \\
&= (U\Sigma V^T)(V\Sigma^T U^T) \\
&= U(\Sigma\Sigma^T)U^T
\end{aligned}$$

$$\begin{aligned}
V &= A^T A \\
&= (U\Sigma V^T)^T (U\Sigma V^T) \\
&= (V\Sigma^T U^T)(U\Sigma V^T) \\
&= V(\Sigma^T \Sigma)V^T
\end{aligned}$$

즉  $u_1, \dots, u_k, \dots, u_m$  는  $\text{range}(A)$  의 직교정규벡터,  $v_1, \dots, v_k, \dots, v_n$  는  $\mathcal{N}(A)^\perp$  의 직교정규벡터이다.

### 1.3.5 특이값 분해에 대한 추가 설명

여기서는 Boyd 교수의 강의노트<sup>1</sup>를 참고하였다. 잠시 편의를 위해  $U$  는  $m \times n$  행렬,  $\Sigma$  는  $n \times n$  행렬이라고 하자. 그러면 특이값 분해는

$$A = U\Sigma V^T = \sum_{i=1}^n \sigma_i u_i v_i^T$$

가 된다. 선형 사상(mapping)  $y = Ax$  는 다음과 같이 분해할 수 있다.

- $x$  를 input direction 들  $v_1, \dots, v_n$  을 따라 계수들을 계산한다.
- $\sigma_i$  는 척도계수

---

<sup>1</sup><https://web.stanford.edu/class/archive/ee/ee263/ee263.1082/notes/ee263coursereader.pdf>

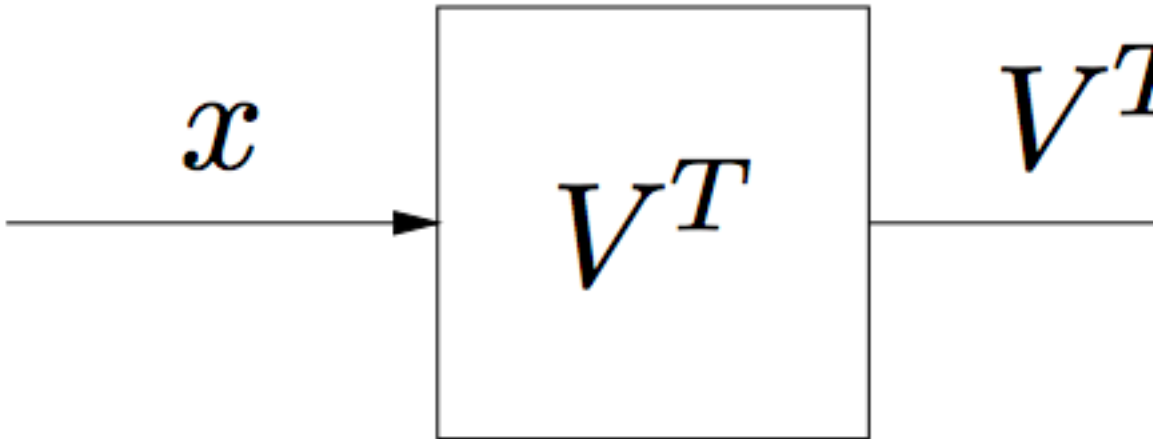


그림 1.1: SVD 그림.

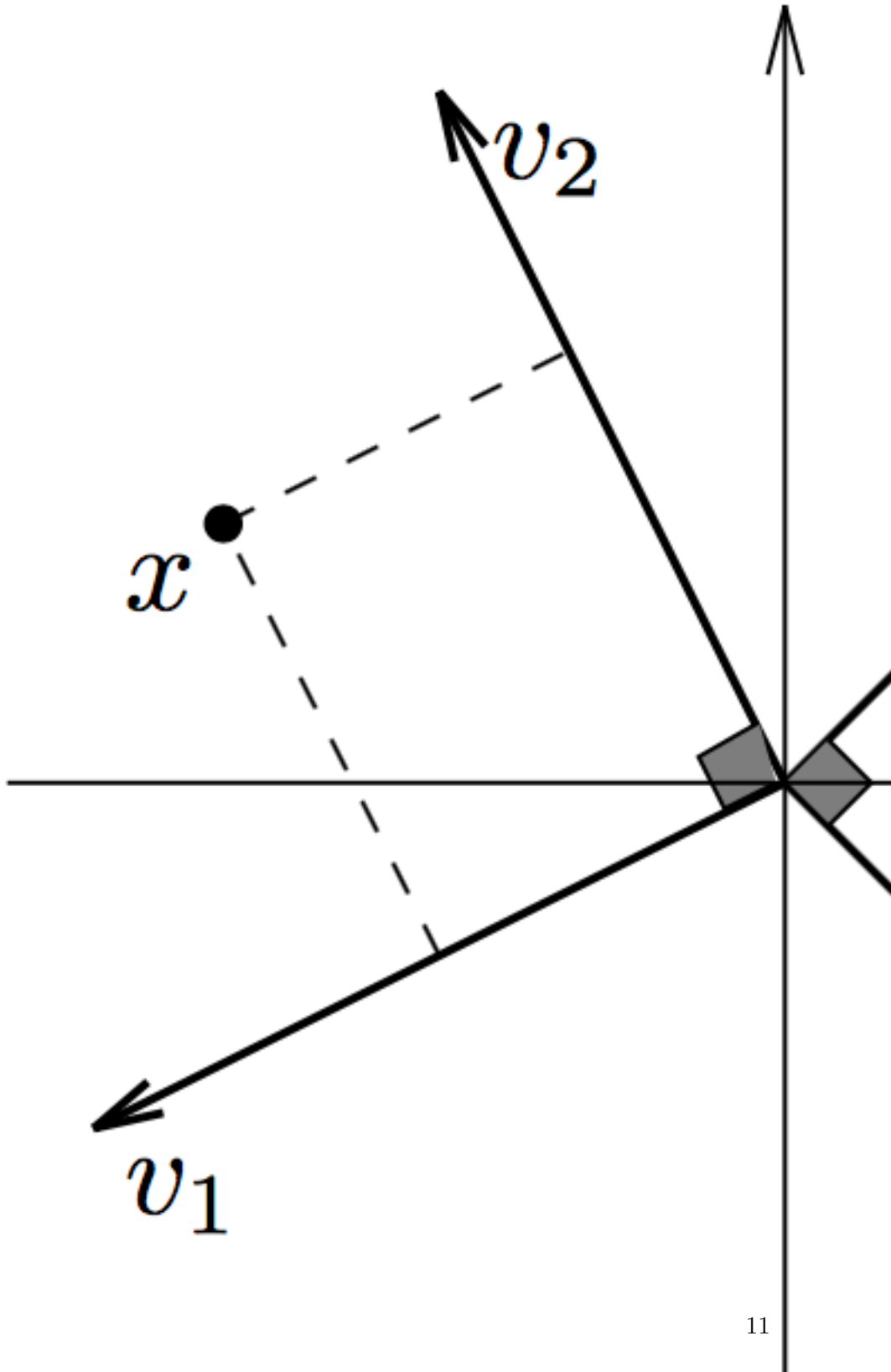
- 이것들을 다시 output directions  $u_1, \dots, u_n$  을 따라 재구성한다.

대칭행렬  $A$  에 대한 고유값 분해와 달라지는 점은 input direction들과 output direction들이 다르다는 것이다.

- $v_1$  는 input direction으로 가장 민감하다.(most sensitive, highest gain)
- $u_1$  은 output direction으로 가장 민감하다.
- $Av_1 = \sigma_1 u_1$  이다.

**Example 1.3.1** (SVD의 기하학적 의미 예).  $A = \mathbb{R}^{2 \times 2}$ 이며  $\Sigma = \text{diag}(1, 0.5)$  인 경우를 생각해보자. 이 경우  $x$  를  $v_1, v_2$  를 따라 풀면  $v_1^T x = 0.5, v_2^T x = 0.6$  즉  $x = 0.5v_1 + 0.6v_2$  이며,  $Ax = (v_1^T x)\sigma_1 u_1 + (v_2^T x)\sigma_2 u_2 = (0.5)(1)u_1 + (0.6)(0.5)u_2$  이다.  $\square$





### 1.3.6 특이값 분해의 기하학적 의미

위키피디아<sup>2</sup>를 참고하자.

## 1.4 기저(basis)

### 1.4.1 Riesz basis

In the Hilbert space  $L_2[0, 1]$ , an unconditional basis is called a **Riesz basis** if it is “almost normalized”. This means that there exist real, positive, non-zero constants  $m$  and  $M$  so that

$$0 < m \leq \|\phi_i\| \leq M < \infty.$$

A Riesz basis is characterized by two Riesz constants  $A$  and  $B$ , so that for all  $f = \sum_i s_i \phi_i \in L_2[0, 1]$ ,

$$A^2 \|f\|^2 \leq \sum_{i \in \mathbb{Z}} s_i^2 \leq B^2 \|f\|^2.$$

(Jensen의 Noise reduction and wavelet thresholding으로부터)

There exists  $\phi_0(x) \in \mathcal{V}_1$  such that  $\{\phi_0(x - k) | k \in \mathbb{Z}\}$  forms a Riesz basis of  $\mathcal{V}_1$ , i.e., there exists  $0 < A \leq B < \infty$  such that

$$A \|c_k\|^2 \leq \left\| \sum_k c_k \phi_0(x - k) \right\|^2 \leq B \|c_k\|^2$$

for all  $\{c_k\} \in l^2$ , where  $A$  and  $B$  do not depend on the  $c_k$

(동익이형 박사논문 47쪽)

### 1.4.2 Radial basis function

Radial function이란 거리에만 의존하는 함수를 의미한다. 어떤 함수에 대한 근사 모델을 radial function의 선형조합으로 표현할 수 있다.

- Gaussian

---

<sup>2</sup>[https://en.wikipedia.org/wiki/Singular\\_value\\_decomposition](https://en.wikipedia.org/wiki/Singular_value_decomposition)

$$\phi(r) = e^{-(\epsilon r)^2}$$

- Multiquadric

$$\phi(r) = \sqrt{1 + (\epsilon r)^2}$$

- Inverse quadratic

$$\phi(r) = \frac{1}{1 + (\epsilon r)^2}$$

- Inverse multiquadric

$$\phi(r) = \frac{1}{\sqrt{1 + (\epsilon r)^2}}$$

## 1.5 공간(space)

이 부분은 전체적으로 (?)의 정의와 내용들을 따라간다.

### 1.5.1 벡터 공간(vector space)

#### 1.5.1.1 완비 벡터 공간(complete vector spaces)

벡터 공간이 유한차원일 때, 공간의 완비성(completeness)은 같은 공간의 다른 larger orthonormal set에 포함되어 있지 않는 orthonormal set을 찾음으로써 증명할 수 있다. 예를 들면 3차원 공간에서는 linear combination이 그 공간의 모든 벡터를 표현할 수 있는 three orthonormal vector의 set을 찾기만 하면 되는 것이다. 그러나 우리가 무한 차원 공간을 고려할 때, 무한한 숫자의 orthonormal vector를 고려하는 것은 쉽지 않다. 사실 무한한 숫자의 vector의 linear combination은 때때로 같은 공간에 포함된 벡터를 표현하는데 충분치 않을 수도 있다. 이러한 모호함을 해결하기 위해 무한 차원 공간에서 완비성을 생각하는 것이다.

**Definition 1.5.1** (벡터의 Cauchy sequence). 벡터의 수열  $\{\mathbf{x}_1, \mathbf{x}_2, \dots\}$ 가 모든  $\epsilon > 0$ 에 대해 적당한 근사 숫자  $N$ 이 존재해 모든  $m, n > N$ 에 대해  $\|\mathbf{x}_m - \mathbf{x}_n\| < \epsilon$ 을 만족한다면 이를 벡터의 **코시 수열 (Cauchy sequence)**이라고 한다. 쉽게 얘기하자면  $\mathbf{x}_m$ 과  $\mathbf{x}_n$ 이  $m, n \rightarrow \infty$  함체 따라 가까아지는 수열을 코시 수열이라 부르는 것이다.  $\square$

**Definition 1.5.2** (코시 수열의 수렴). 벡터의 무한 수열  $\{\mathbf{x}_1, \mathbf{x}_2, \dots\}$ 가 있을 때, 만약  $\mathbf{x}$ 가 존재해  $\|\mathbf{x}_n - \mathbf{x}\| \rightarrow 0$ 을 만족한다면 이 수열이 수렴 (convergent)한다고 한다.  $\square$

## 1.5.2 위상공간(topological space)

## 1.5.3 분해 가능 공간(separable space)

Countable dense subset을 포함하는 위상공간(topological space)을 **분해 가능 공간 (separable space)**이라고 한다.

## 1.5.4 노름 공간(normed space)

**Definition 1.5.3** (노름 공간). 원소들에 일종의 ‘길이’ 또는 ‘크기’가 부여된 벡터 공간  $V$ 을 **노름 공간 (normed space)**라고 한다.  $\square$

노름 공간에 있는 모든 원소들은 항상 유한한 노름을 갖아야 한다. 만약 무한한 노름을 갖는다면 그것은 공간이 될 수 없다. 노름  $\|\cdot\|$ 은 항상 계량(metric)을 유도한다. 계량은  $\mathcal{F}$ 에서의 거리 개념이다. 즉  $d(f, g) = \|f - g\|_{\mathcal{F}}$ 이다. 이것은  $\mathcal{F}$ 가 어떤 위상 구조를 갖고있다는 뜻도 되며 우리는  $\mathcal{F}$ 의 원소들의 수열들이 정의된 거리에서 수렴하는지, 또는 연속 등에 대해서 생각할 수 있게 한다.

## 1.5.5 바나흐 공간(Banach space)

**Definition 1.5.4** (힐버트 공간). Normed space가 complete일 경우 이를 **바나흐 공간 (Banach space)**라고 부른다.  $\square$

유한차원에서 노름공간은 항상 완비(complete)이다. 그러나 무한차원에서 노름공간은 완비일 수도 있고, 아닐 수도 있다.

**Definition 1.5.5** (원-힐버트 공간). Inner product spaces들 중 완비가 아닌 공간들을 원-힐버트 공간(pre-Hilbert space)라고 부른다.  $\square$

### 1.5.6 힐버트 공간(Hilbert space)

**Definition 1.5.6** (힐버트 공간). A complete normed space endowed with inner product is called a **Hilbert space**  $\square$

다음 그림은 (?)에 있는 그림으로 각 공간들 간의 사이의 관계를 이해하는 데 도움이 된다.

힐버트 공간의 대표적인 예로,  $L^p$  공간이 있다.

$$\|f\|_p = \left( \int_S |f|^p d\mu \right)^{1/p} < \infty$$

박창이 교수님 외 책에서, 통계학에서 힐버트 공간을 고려하는 이유를 설명하였는데, 그 이유는 유한차원에서 성립하는 선형대수를 적용할 수 있기 때문이라고 한다. 특히 힐버트 공간은 완비(complete)된 공간으로써 알고리즘의 수렴성을 보장하며, 내적이 존재하므로써 직교성이나 정사영(projection)을 구할 수도 있다.

### 1.5.7 Holder 공간(Holder space)

이 공간의 정의는 (?)를 참고하였다. 어떤 정수  $m$ 과  $\delta \in (0, 1)$ 에 대해  $s = m + \delta$ 을 정의하자. **Holder 공간(Holder space)**(정확하게는  $H^s$ {o}lder space)은 bounded  $m$ 차 derivative를 갖는, 즉 모든  $u, t$ 에 대해  $|f^m(u) - f^m(t)| \leq |u - t|^\delta$ 인 bounded function들의 집합이다.

### 1.5.8 Sobolev 공간(Sobolev space)

이 공간의 정의는 (?)를 참고하였다. 이 책에서 Sobolev 공간의 정의는 smooth function들의 집합이라는 것이다.  $D^j f$ 를  $f$ 의  $j$ 번째 weak derivative라고 하자. 먼저 weak differentiability에 대해 알아보자.  $f$ 가 모든 bounded interval에서 적분 가능하다고 하자. 만약 다음과 같은 함수  $f'$ 가 존재해 모든 bounded interval에서 적분 가능하다면, 즉 모든  $x \leq y$

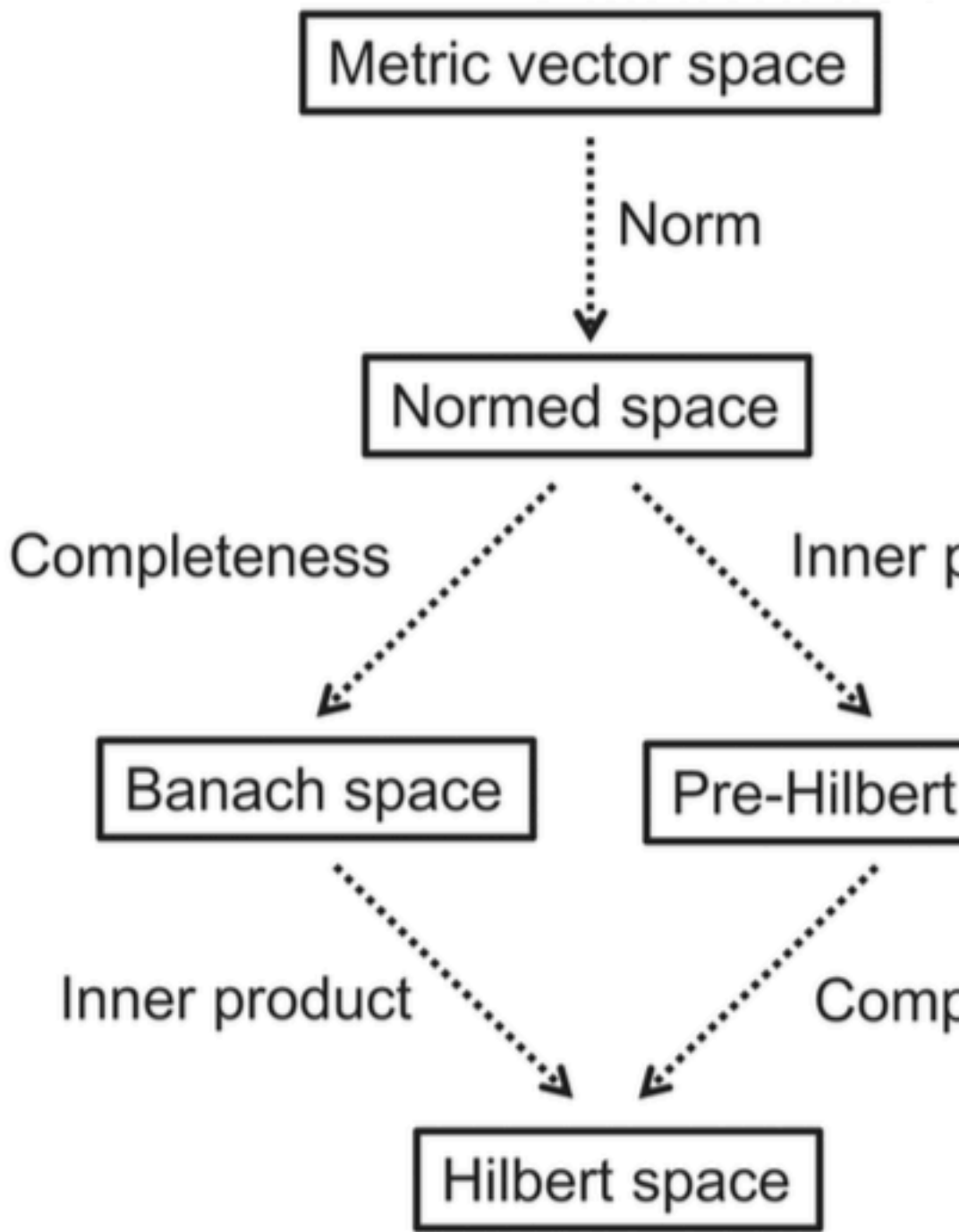


그림 1.3: Metric vector space로부터 힐버트 공간까지의 관계.

에 대해

$$\int_x^y f'(s)ds = f(y) - f(x)$$

이라면 이를 **weakly differentiable**이라 부른다.

**Definition 1.5.7** (Sobolev 공간).  **$m$ 차 Sobolev space**는 다음과 같다.

$$W(m) = \{f \in L^2(0, 1) : D^m f \in L^2(0, 1)\}$$

반지를  $c$ 를 갖는  $m$ 차 Sobolev space는 다음과 같다.

$$W(m, c) = \{f \in W(m), \|D^m f\|^2 \leq c^2\}$$

주기를 갖는 Sobolev class (**periodic Sobolev class**)는 다음과 같다.

$$\tilde{W}(m, c) = \{f \in W(m, c) : D^j f(0) = D^j f(1), j = 0, \dots, m-1\}.$$

□

### 1.5.9 Besov 공간(Besov space)

이 공간의 정의 또한 (?)의 9장을 참고하였다. Wavelet threshold estimator는 Besov space에서 good optimality property를 갖는다. 다음과 같이  $\Delta_h^{(r)} f(x)$ 를 정의하자.

$$\Delta_h^{(r)} f(x) = \sum_{k=0}^r \binom{r}{k} (-1)^k f(x + kh).$$

그러면  $\Delta_h^{(0)} f(x) = f(x)$ 이고

$$\Delta_h^{(r)} f(x) = \Delta_h^{(r-1)} f(x + h) - \Delta_h^{(r-1)} f(x).$$

이다. 이제  $w_{r,p}(f; t)$ 를 다음과 같이 정의하자.

$$w_{r,p}(f; t) = \sup_{|h| \leq t} \|\Delta_h^{(r)} f\|_p.$$

이 때  $\|g\|_p = \{\int |g(x)|^p dx\}^{1/p}$ 이다.  $(p, g, \zeta)$ 가 주어졌을 때,  $r$ 이 존재해  $r-1 \leq \zeta \leq r$ 을 만족한다면, **Besov seminorm**  $|f|_{p,q}^\zeta$ 는

$$|f|_{p,q}^\zeta = [\int_0^\infty (h^{-\zeta} w_{r,p}(f; h))^q \frac{dh}{h}]^{1/q}$$

로 정의된다.  $q = \infty$ 일 때는 다음과 같이 정의한다.

$$|f|_{p,\infty}^\zeta = \sup_{0 < h < 1} \frac{w_{r,p}(f; h)}{h^\zeta}$$

**Definition 1.5.8** (Besov 공간). **Besov 공간(Besov space)**  $B_{p,q}^\xi(c)$ 는  $[0, 1]$ 에서  $\mathbb{R}$ 로 가는 함수  $f$ 들 중  $\int |f|^p < \infty$ 이고  $|f|_{p,q}^\xi \leq c$ 인  $f$ 들의 집합입니다.  $\square$

Sobolev space  $W(m)$ 은 Besov ball  $B_{2,2}^m$ 에 대응된다. Generalized Sobolev space  $W_p(m)$ 은  $m$ 차 미분에서  $L^p$  노름을 사용하는 데 이것은 거의 Besov space와 비슷하다. 사실  $B_{p,1}^m \subset W_p(m) \subset B_{p,\infty}^m$ 이다. Holder space는  $B_{\infty,\infty}^{m+\delta}$ 와 같다.  $T$ 를 bounded variation을 갖는 함수들을 포함하는 집합이라고 할 때,  $B_{1,1}^1 \subset T \subset B_{1,\infty}^1$ 을 만족한다. 즉 Besov space는 넓은 범위의 함수 공간들을 포함한다. 이에 수렴하는 수열 (convergent sequence)와 코시 수열 (Cauchy sequence)에 대해 생각해 볼 수 있다.

### 1.5.10 Skorohod 공간(Skorohod space)

$D[a, b]$ 을 right continuous하고 left limit을 갖는 함수들  $z : [a, b] \rightarrow \mathbb{R}$ 의 집합을 **Skorohod 공간(Skorohod space)**이라고 한다. (?) 257쪽에 나와있다.

#### 1.5.10.1 웨이블릿 expansion의 계수들로 이해하는 Besov 공간(Besov spaces in terms of the coefficients of the wavelet expansion)

### 1.5.11 재생 커널 힐버트 공간(reproducing kernel Hilbert space)

재생 커널 힐버트 공간(reproducing kernel Hilbert space)이 주어지면 이에 대응되는 유일한 재생 커널이 존재한다. 역으로, 양정치 커널  $K$ 가 주어지면  $K$ 를 재생 커널로 갖는 유일한 재생 커널 힐버트 공간  $\mathcal{H}$ 가 존재한다.

## 1.6 거리(distance)

군집분석 방법에서는 관측값들의 거리를 이용해 군집을 나눌 때 사용된다.

- 유클리드 거리(Euclidean distance):  $d(x, y) = (\sum_{i=1}^p (x_i - y_i)^2)^{1/2}$
- 민콥스키 거리(Minkowski distance):  $d(x, y) = (\sum_{i=1}^p (x_i - y_i)^m)^{1/m}$
- 맨하탄 거리(Manhattan distance):  $d(x, y) = \sum_{i=1}^p |x_i - y_i|$



- 표준화 거리 (standardized distance):  $d(x, y) = (\sum_{i=1}^p (x_i - y_i)^2 / s_i^2)^{1/2}$ , 여기서  $s_i$ 는  $i$ 번째 변수에 대한 표준편차
- 마할라노비스 거리 (Mahalanobis distance):  $d(x, y) = (x - y)^T \Sigma^{-1} (x - y)$ , 여기서  $\Sigma$ 는 공분산행렬
- 체비셰프 거리 (Chebychev distance):  $d(x, y) = \max_{i=1, \dots, p} |x_i - y_i|$

다음 거리들은 유클리드 거리와 더불어 공간통계에서 많이 쓰이는 것들이다.

- chordal distance (현 거리? 잘 모르겠음)
- geodesic distance

## 1.7 계량 (metric)

계량 (metric)은 두 집합 사이의 모든 원소 짝의 거리를 지정해주는 함수이다.

## 1.8 실함수의 수열들 (sequences of real functions)

확률변수 (random variable)의 수열이 함수의 수열처럼 여겨질 수 있다는 사실에서 함수의 수열의 성질을 이해하는 것은 중요하다. 일반적인 수열의 수렴에 관한 성질들과 비교했을 때 함수의 수열들의 수렴에 관한 성질에서는 여러 개의 정의로 다뤄질 수 있다는 점에서 차이가 있다. 가장 널리 알려진 함수의 수렴은 **점별수렴 (pointwise convergence)**과 **균등수렴 (uniform convergence)**이 있다.

**Definition 1.8.1** (점별수렴).  $\{f_n(x)\}_{n=1}^\infty$ 를 실함수의 수열이라고 하자. 그러면 모든  $x \in \mathbb{R}$ 에 대해

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

를 만족하는 실함수  $f$ 가 존재할 때 수열  $\{f_n(x)\}_{n=1}^\infty$ 이  $f$ 에 **점별수렴 (pointwise convergence)**한다고 한다. 여기서는  $n \rightarrow \infty$ 일 때

$$f_n \xrightarrow{pw} f$$

로 표현하기로 한다. □

점별수렴보다 더 강한 조건으로  $x \in \mathbb{R}$ 에 상관 없이 모든 지점에서 동시에  $n \rightarrow \infty$ 일 때  $f_n(x) \rightarrow f(x)$ 이길 요구할 수도 있다. 이 때 사용되는 정의가 균등수렴이다.

**Definition 1.8.2** (균등수렴).  $\{f_n(x)\}_{n=1}^\infty$ 를 실함수의 수열이라고 하자. 그러면 모든  $\epsilon > 0$ 에 대해

$$|f_n(x) - f(x)| < \epsilon \forall n \geq n_\epsilon \text{ and } x \in \mathbb{R}$$

을 만족시키는 정수  $n_\epsilon$ 이 존재할 때 수열  $\{f_n(x)\}_{n=1}^\infty$ 이  $f$ 에 **균등수렴 (uniform convergence)**한다고 한다. 여기서는  $n \rightarrow \infty$ 일 때

$$f_n \xrightarrow{u} f$$

로 표현하기로 한다. □

(두 수열의 차이점 설명)

### 1.8.1 수렴하는 수열 (convergent sequence)

**Definition 1.8.3** ((함수해석학에서) 수렴하는 수열). 노름 벡터 공간  $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ 의 원소들로 이루어진 수열  $\{f_n\}_{n=1}^\infty$ 가 모든  $\epsilon > 0$ 에 대해  $N = N(\epsilon) \in \mathbb{N}$ 이 존재해 모든  $n \geq N$ 에 대해  $\|f_n - f\|_{\mathcal{F}} < \epsilon$ 을 만족할 때, 우리는 수열이  $f \in \mathcal{F}$ 에 **수렴 (converge)**한다고 한다. □

### 1.8.2 코시 수열 (Cauchy sequence)

**Definition 1.8.4** ((함수해석학에서) 수렴하는 수열). 노름 벡터 공간  $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ 의 원소들로 이루어진 수열  $\{f_n\}_{n=1}^\infty$ 가 모든  $\epsilon > 0$ 에 대해  $N = N(\epsilon) \in \mathbb{N}$ 이 존재해 모든  $m, n \geq N$ 에 대해  $\|f_m - f_n\|_{\mathcal{F}} < \epsilon$ 을 만족할 때, 수열  $\{f_n\}_{n=1}^\infty$ 를 **코시 수열 (Cauchy sequence)**이라고 한다. □

**Example 1.8.1** (유리수의 장). 절대값  $\|\cdot\|$ 을 노름으로 갖는 유리수  $\mathbb{Q}$ 의 장(field)은 노름 벡터 공간이다. 여기서 수열  $1, 1.4, 1.41, \dots$ 는  $\mathbb{Q}$ 에서의 코시 수열이나 이것의 극한  $\sqrt{2} \notin \mathbb{Q}$ 이므로 수렴하는 수열은 아니다. □

**Example 1.8.2** ( $C[0,1]$ ).  $C[0,1]$ 을 구간  $[0,1]$ 에서 bounded continuous인 함수들의 집합이라고 하자. 다음과 같은 노름  $\|f\| = (\int_0^1 |f(x)|^2 dx)^{1/2}$ 을 생각할 때 함수들의 수열

$\{f_n\}$  은  $[0, \frac{1}{2} - \frac{1}{2n}]$  에서 0 이고  $[\frac{1}{2} + \frac{1}{2n}, 1]$  에서 1 이기 때문에 코시 수열이나 연속 극한 (continuous limit) 은 존재하지 않는다.  $\square$

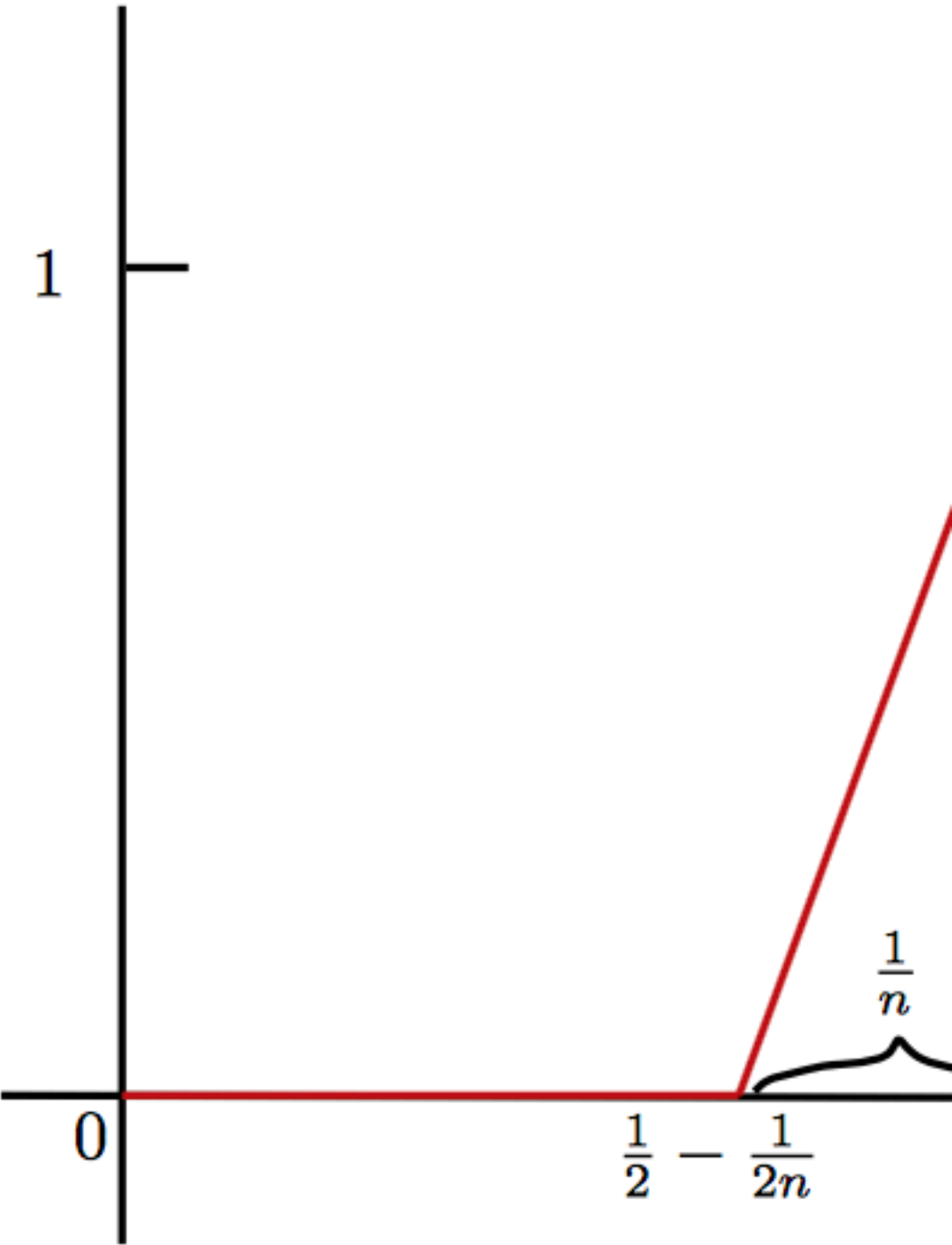


그림 1.4: L2 norm에서 연속 극한을 갖지 않는 연속함수들의 코시 수열의 예.

## 2 기초 확률론

이 장에서는 앞으로 다룰 내용을 이해하기 위해 필요한 기본적인 확률 개념을 정리하였다. 대학원 과정의 확률론을 다룬 유명한 책들로는 (?), (?) 그리고 (?)이 있다. 그 밖에 본인이 추천하는 책들은 다음과 같다. (?)는 최근에 나온 대학원 확률론 입문서 교재로써 비교적 내용이 자세하다. (?)는 삽화가 많고 저자가 연습문제의 답을 웹에 올려놓았다. (?)과 (?)는 통계학자의 입장에서 필요한 확률론 지식을 비교적 쉽게 서술하였다. 여기서는 앞서 언급한 모든 책들을 참고할 것이다.

### 2.1 표본공간과 사건(sample space and events)

통계학은 무작위(random) 또는 확률적(stochastic) 실험(experiment), 즉 어떤 결과가 나올지 미리 확실히 예측할 수 없는 실험들에 초점을 맞춘다.

**Definition 2.1.1** (표본공간). 어떤 무작위 실험의 **표본공간(sample space)**  $\Omega$ 는 그 실험에서 나올 수 있는 모든 결과들의 집합이다.  $\square$

**Example 2.1.1** (동전 던지기 실험). 동전을 두 번 던지는 실험에서  $\Omega = \{HH, HT, TH, TT\}$  이다. 이러한 표본공간을 (**유한**) **이산 표본공간(finite discrete sample space)**이라고 한다.  $\square$

**Example 2.1.2** (동전 계속 던지기 실험). 이번에는 동전의 뒷 면이 나올때까지 동전을 계속해서 던지는 실험에 대해 살펴보자. 그러면

$$\{T, HT, HHT, HHHT, \dots, \{HHH \dots\}\}$$

와 같은 결과들의 수열을 얻을 수 있다. 이를 만약 동전을 던진 횟수로 정리한다면

$$\{1, 2, 3, \dots, \infty\}$$

로 볼 수 있다. 이러한 표본공간을 (무한) 이산 표본공간 (infinite discrete sample space)라고 한다.  $\square$

**Example 2.1.3** (지하철 도착 시간). 우리가 지하철을 기다리고 있다고 가정해보자. 지하철은  $T$  시간마다 한 번씩 도착한다. 그러면 우리가 기다리는 시간에 대한 표본공간은

$$[0, T] = \{t : 0 \leq t \leq T\}$$

이다. 이러한 표본 공간은 연속 표본공간 (continuous sample space)라고 한다.  $\square$

**Definition 2.1.2** (사건). 사건 (event)란 표본공간  $\Omega$ 의 임의의 부분집합 (subset)을 의미한다.  $\square$

**Example 2.1.4** (사건의 예). - 앞서 동전을 두 번 던지는 실험에서 앞면이 하나만 나올 사건을  $A$ 라고 하면  $A = \{HT, TH\}$ 이다.

- 앞서 동전을 두 번 던지는 실험에서 적어도 한 번 앞면이 나올 사건을  $B$ 라고 하면  $A = \{HH, HT, TH\}$ 이다.

$\square$

## 2.2 시그마-체 (sigma-field)

앞서 표본공간  $\Omega$ 의 임의의 부분집합인 사건을 생각했는데, 그러면 이 사건들의 집합  $\mathcal{F}$ 에 대해서도 생각해 볼 수 있을 것이다. 그리고 사건들의 집합이 가져야 할 바람직한 성질들을 잘 정의하기 위해 시그마-체라는 개념을 도입한다.

**Definition 2.2.1** (대수 (체)). 어떤 집합 (set)  $\Omega$ 의 non-empty collection (즉  $\Omega$ 의 subset들의 모임)을  $\mathcal{F}$ 라고 하자. 그러면  $\mathcal{F}$ 가

1.  $\Omega \in \mathcal{F}$  (또는  $\emptyset \in \mathcal{F}$ )
2.  $A \in \mathcal{F}$ 이면  $A^C \in \mathcal{F}$ ,
3.  $A, B \in \mathcal{F}$ 이면  $A \cup B \in \mathcal{F}$

를 만족할 때  $\mathcal{F}$ 를 대수 (algebra) 또는 체 (field)라고 부른다.  $\square$

시그마-체는 앞선 대수의 정의에서 두 번째 조건이 조금 바뀐 것이다.

**Definition 2.2.2** (시그마-체). 어떤 집합(set)  $\Omega$ 의 non-empty collection을  $\mathcal{F}$ 라고 할 때,  $\mathcal{F}$ 가

1.  $\Omega \in \mathcal{F}$  (또는  $\emptyset \in \mathcal{F}$ )
2.  $A \in \mathcal{F}$ 이면  $A^C \in \mathcal{F}$ ,
3.  $A_1, A_2, \dots \in \mathcal{F}$  이면  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

를 만족할 때  $\mathcal{F}$ 를 시그마-대수 (sigma-algebra) 또는 시그마-체 (sigma-field)라고 부른다.  $\square$

다음은 체와 시그마-체에 대한 간단한 사실들이다.

**Corollary 2.2.3** (시그마-체에 대한 사실들). 1. 모든 체는 *finite union*에 대해 닫혀있다. 또한 같은 논리를 적용해 *finite intersection*에 대해서도 닫혀있다.

2. 모든 시그마-체  $\mathcal{F}$ 는 *countable intersection*에 대해서도 닫혀있다. 즉,

$$A_1, A_2, \dots \in \mathcal{F} \text{ 이면 } \bigcap_{i=1}^{\infty} A_i = \left( \bigcap_{i=1}^{\infty} A_i^C \right)^C \in \mathcal{F}.$$

물론 모든  $A_1^C, A_2^C, \dots$  또한  $A_1^C, A_2^C, \dots \in \mathcal{F}$  이다.

3.  $\mathcal{F}$ 가 *non-void*일 경우에는 모든 체 또는 시그마-체가  $A$ 를 포함하고 있으면  $A^C$  또한 포함하고 있기 때문에  $\Omega = A \cup A^C$  와  $\emptyset = \Omega^C$  또한  $\mathcal{F}$ 에 포함되어 있다. 따라서 첫 번째 조건을 생략해도 된다.

$\square$

**Example 2.2.1** (시그마-체의 예). - 어떤 집합  $\Omega$ 에 대해,  $\{\emptyset, \Omega\}$ 는 시그마-체가 된다. 이 시그마-체는  $\Omega$ 의 부분집합으로 만들 수 있는 가장 작은 시그마-체이다.

- $\Omega$ 의 멱집합(power set, 어떤 집합의 모든 부분집합을 모은 집합) 또한 시그마-체이며 이는  $\Omega$ 의 부분집합으로 만들 수 있는 가장 큰 시그마-체이다.
- $A \in \Omega$ 일 때 collection  $\{\emptyset, A, A^C, \Omega\}$  또한 간단히 만들 수 있는 시그마-체의 예다.

$\square$

**Example 2.2.2** (체이나 시그마-체가 아닌 예). 다음은  $\mathcal{F}$ 가 체이나 시그마-체가 아닌 예이다.  $\Omega = (0, 1]$ 이고,  $\mathcal{F}$ 는  $\emptyset$ 과 모든

$$(a, b], \quad a, b \in \mathbb{Q}, a, b \in [0, 1], a < b$$

와  $(a, b]$ 의 모든 finite union을 포함한다고 하자. 그리고  $[z]$ 를  $z$ 와 가장 가까운 정수로 반올림해주는 연산자라고 하자. 그러면 정의에 의해  $\mathcal{F}$ 는 체가 된다. 그러나  $A_n = (a_n, 1]$ ,  $a_n = \frac{10^n}{[10^n \pi]}$  라고 하면

$$A_n \in \mathcal{F} \text{ 이나 } \bigcup_{n=1}^{\infty} A_n = (\pi, 1] \notin \mathcal{F}$$

이다. 따라서  $\mathcal{F}$ 는 시그마-체가 아니다.  $\square$

**Example 2.2.3** (표본공간이 셀 수 있는 집합이면 멱집합이 사건들의 집합). 표본공간  $\Omega$ 가 셀 수 있는 집합, 예를 들면  $\{0, 1, 2, \dots\}$ 라고 가정하자. 그리고 이 때 사건들의 집합  $\mathcal{F}$ 가 모든 singleton  $\omega_i, i = 1, 2, \dots$ 들을 포함하는 시그마-체가 되길 원한다고 가정하자. 그러면  $\Omega$ 의 모든 부분집합  $E$ 는  $\bigcup_{i=1}^{\infty} \omega_i$ 로 만들 수 있다. 즉 singleton들의 countable union으로 만들 수 있다. 그리고 countable union에 대해 시그마-체가 닫혀있기 때문에,  $\mathcal{F}$ 가  $\Omega$ 의 어떤 부분집합  $E$ 들을 모두 포함한다는 결론에 이른다. 즉, **표본공간이 셀 수 있는 집합이면, 우리는 항상 멱집합을 사건들의 집합으로 써야 한다.**  $\square$

## 2.3 생성기들 (generators)

시그마-체에 대해 좀 더 자세히 살펴보기 위해, **생성기 (generator)**에 대해 알아보자. 표본공간  $\Omega$ 의 subset들의 collection  $\mathcal{A}$ 가 있다고 하자. 그러면 멱집합은 항상 시그마-체이기 때문에,  $\mathcal{A}$ 를 포함하는 시그마-체가 적어도 한 개 이상 있을 것이다.  $\mathcal{F}^*$ 를  $\mathcal{A}$ 를 포함하는 모든 시그마-체의 모임, 즉

$$\mathcal{F}^* = \{\sigma\text{-algebras } \supset \mathcal{A}\}$$

라고 하자. 여기서  $\mathcal{A}$ 를 포함하는 **가장 작은** 시그마-체를 생각해보자. 즉

$$\mathcal{F} = \sigma(\mathcal{A}) = \bigcap_{\{\mathcal{F}' \text{ } \sigma\text{-algebra } | \mathcal{A} \subset \mathcal{F}'\}} \mathcal{F}' = \bigcap_{\mathcal{G} \in \mathcal{F}^*} \mathcal{G}$$

인  $\mathcal{F}$ 가 존재하고 이를  $\mathcal{A}$ 로부터 **생성된 시그마-체 (sigma algebra generated by  $\mathcal{A}$ )**라고 부른다.

**Example 2.3.1** (생성기들). - 만약  $\mathcal{A} = A$ , 즉  $\mathcal{A}$ 가 single set일 경우  $\sigma(\mathcal{A}) = \{\emptyset, A, A^C, \Omega\}$ 이다.

- 만약  $\mathcal{A}$ 가 시그마-체일 경우,  $\sigma(\mathcal{A}) = \mathcal{A}$ 다.

$\square$



## 2.4 확률공간 (probability space)

**Definition 2.4.1** (가측공간). 표본공간  $\Omega$ 와 이와 연관된 시그마-체  $\mathcal{A}$ 를 묶어  $(\Omega, \mathcal{A})$ 를 가측공간 (measurable space)이라고 한다.  $\square$

**Definition 2.4.2** (확률측도). 가측공간  $(\Omega, \mathcal{A})$ 가 주어졌을 때 **확률측도** (probability measure)  $P$ 는  $P : \mathcal{A} \rightarrow [0, 1]$ 인 함수로

1.  $P(\emptyset) = 0$  and  $P(\Omega) = 1$
2. 어떤  $A \in \mathcal{A}$ 에 대해  $P(A) \geq 0$
3. (가산가법성 (countable additivity)):  $\{A_n, n \geq 1\}$ 이 disjoint 라고 하면  $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ .

을 만족한다. 그리고  $(\Omega, \mathcal{A}, P)$ 를 묶어 **확률공간** (probability space)라고 한다.  $\square$

이 확률측도는  $\mathcal{A}$ 가 시그마-체일 때 뿐 아니라 그냥 체 일때도 위 세 가지를 만족하면 정의할 수 있다.

## 2.5 보렐 시그마-체 (Borel sigma field)

이제  $\Omega$ 가 비가산집합 (uncountable set) 일 때 시그마-체에 대해 살펴보자. 비가산집합의 대표적인 예로  $\mathbb{R}$ 이 있으니  $\Omega = \mathbb{R}$ 이라 놓고 전개하기로 한다. 앞서 얘기했듯이 시그마-체의 크기는 우리가 고려하고 싶은 모든 사건들과 그 사건들의 countable union, intersection을 적절히 잘 포함하는 정도여야 한다. 가장 쉽게 만들 수 있는 것은  $\mathcal{F}$ 가 모든 countable subset  $E$ 를 포함하게끔 만드는 것이다. 그러나 이 시그마-체는 충분히 크지 않다. 예를 들어  $\Omega = [0, 1]$  일 경우, 앞서 말한 대로  $\mathcal{F}$ 를 만들면  $[0, 0.5]$  같은 사건은 countable이나 co-countable이 아니므로  $\mathcal{F}$ 에 포함이 되지 않는 것이다.

즉 우리는  $\Omega$ 의 모든 interval들을 포함하는 시그마-체를 만들고 싶어한다. 예를 들면,  $\Omega = [0, 1]$  일 때

$$(a, b) \in \mathcal{F}, \quad (0 \leq a < b \leq 1),$$

$$P((a, b)) = b - a, \quad (0 \leq a < b \leq 1)$$

이 되길 원하는 것이다. 가장 간단한 방법으로, 멱집합을  $\mathcal{F}$ 로 나용할 수 있다. 그러나 이  $\mathcal{F}$ 은 너무 크다.  $\mathcal{F}$ 가 너무 클 경우, 확률측도가 잘 construct되지 않는 경우가 생길 수 있다고 한다.

### 2.5.1 연속 표본공간에서 시그마-체로 멱집합을 쓰지 않는 이유 (no uniform probability of power set on continuous sample space)

멱집합이 시그마-체로 적합하지 않은 이유로  $([0, 1], 2^{[0,1]})$ 에서 균등확률이 존재하지 않음을 보일 것이다.  $P$ 를  $([0, 1], 2^{[0,1]})$ 에서 균등확률의 한 후보라고 놓자. 우리는  $P$ 가

$$P\{[a, b]\} = P\{(a, b)\} = P\{[a, b)\} = P\{(a, b]\} = b - a, \quad \text{for any } [a, b] \subseteq [0, 1]$$

을 만족하길 원한다. 또한 특별히

$$P\{a\} = 0, \quad \text{for every } 0 \leq a \leq 1$$

이다. 그리고  $P$ 는 확률의 공리(the axioms of probability) 중 하나인 가산가법성(countable additivity)을 만족시켜야 한다. 즉  $0 \leq a_1 < b_1 < \dots < a_n < b_n < \dots \leq 1$ 이면  $P$ 는

$$P\left\{\bigcup_{i=1}^{\infty} [a_i, b_i]\right\} = \sum_{i=1}^{\infty} P\{[a_i, b_i]\} = \sum_{i=1}^{\infty} (b_i - a_i)$$

를 만족해야 한다.

또한  $P$ 는 이동불변(shift invariant) 성질을 가져야 한다. 즉, 확률은 interval의 length에만 영향을 받아야 한다.

$$P\{[r, 1/4 + r]\} = \frac{1}{4}, \quad \text{for every } 0 < r \leq 3/4.$$

그런데 한 가지 문제가 생기는데  $3/4 < r < 1$ 이면  $[r, 1/4 + r]$ 이  $[0, 1]$ 의 부분집합이 되지 않는다. 이를 해결하기 위해 “wrapping around”라는 방법을 이용한다. 만약 “wrapping around”를  $\oplus$ 로 나타낸다면

$$[0, 1/4] \oplus r = \begin{cases} [r, 1/4 + r] & \text{if } 0 < r \leq 3/4 \\ [0, 1/4 + r - 1] \cup [r, 1] & \text{if } 3/4 < r < 1 \end{cases}$$

로 정의하는 것이다. 그러면  $A \subseteq [0, 1]$ 이라고 할 때  $A$ 를  $r$ ( $0 < r < 1$ )만큼 이동하는 것을

$$A \oplus r = \{a + r : a \in A, a + r \leq 1\} \cup \{a + r - 1 : a \in A, a + r > 1\}$$

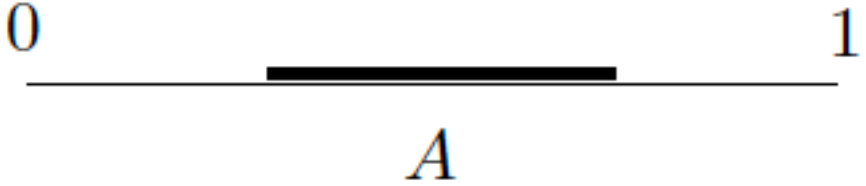


그림 2.1: Shift invariance.

로 정의할 수 있다.

“wrapping around”를 이용해  $A$ 를  $r$ 만큼 이동해도 길이가 보존되기 때문에, 확률 또한

$$P\{A \oplus r\} = P\{A\}, \quad \text{for any } 0 < r < 1$$

이 될 것이라 추론할 수 있다.

이제 모든  $A \in 2^{[0,1]}$ 에 대해 균등확률이 존재하지 않음을 보이기 위해 동치관계 (equivalence relation)라는 것에 대해 정의할 것이다.  $x$ 와  $y$  ( $x, y \in [0, 1]$ )는  $y - x \in \mathbb{Q}$ 를 만족할 경우 동치관계라 정의하고  $x \sim y$ 로 표시한다. 예를 들면

$$\frac{1}{2} \sim \frac{1}{4}, \frac{1}{3} \sim \frac{1}{\pi}, \frac{1}{\pi} - \frac{1}{4} \sim \frac{1}{\pi} + \frac{1}{2}$$

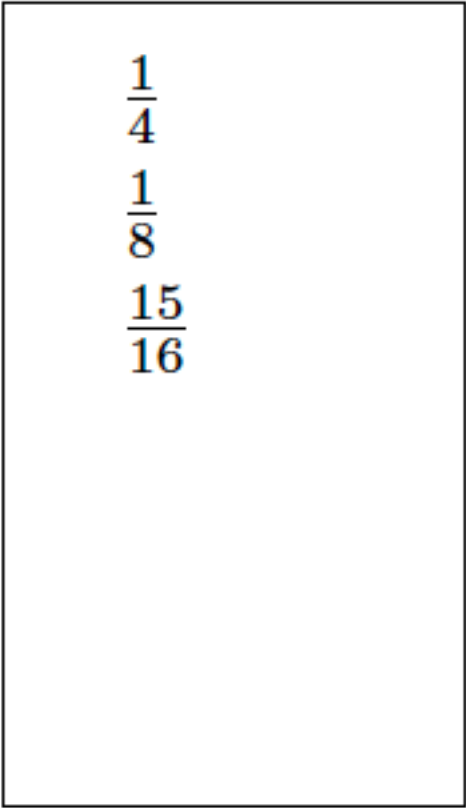
인 것이다.

이 동치관계는  $[0, 1]$ 을 다음과 같이 분리 (disjoint) 합집합들로 표현할 수 있다.

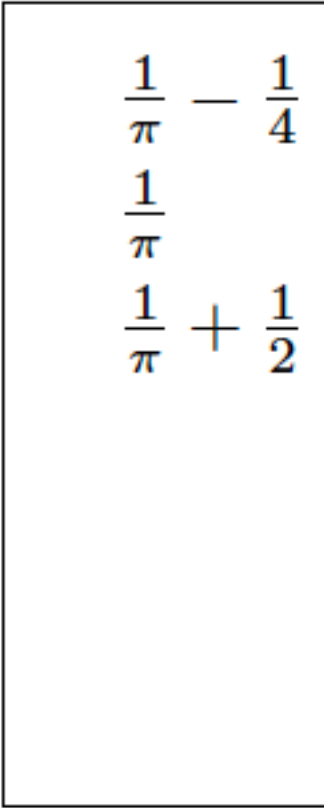
$$[0, 1] = \mathbb{Q}_1 \cup \left\{ \bigcup_{x \in [0,1] \setminus \mathbb{Q}_1} \{(\mathbb{Q} + x) \cap [0, 1]\} \right\} = \mathbb{Q}_1 \cup \left\{ \bigcup_{x \in [0,1] \setminus \mathbb{Q}_1} \{(\mathbb{Q} + x) \oplus x\} \right\}.$$

$H$ 를 선택공리 (the Axiom of Choice)에 의해  $[0, 1]$ 의 모든 동치관계에서 원소를 한 개씩 잘 뽑아서 만든  $[0, 1]$ 의 부분집합이라고 하자. 편의상  $0 \notin H$ 라고 하자. 그러면  $(0, 1]$ 을

$$(0, 1] = \bigcup_{r \in \mathbb{Q}_1, r \neq 1} \{H \oplus r\} \quad \text{with } \{H \oplus r_i\} \cap \{H \oplus r_j\} = \emptyset \text{ for all } i \neq j$$



$$\mathbb{Q}_1$$



$$(\mathbb{Q} + \frac{1}{\pi}) \cap [0,1]$$

그림 2.2: Collection of disjoint unions on  $[0,1]$

로 표현할 수 있다. 그러면

$$1 = P\{(0, 1]\} = P\left\{\bigcup_{r \in \mathbb{Q}_1, r \neq 1} \{H \oplus r\}\right\} = \sum_{r \in \mathbb{Q}_1, r \neq 1} P\{H \oplus r\} = \sum_{r \in \mathbb{Q}_1, r \neq 1} P\{H\}$$

가 된다. 만약 우리가  $p = P\{H\}$ 로 확률을 부여하고자 한다면

1.  $p = 0$ 일 때에는  $1 = \sum_{r \in \mathbb{Q}_1, r \neq 1} P\{H\} = \sum_{r \in \mathbb{Q}_1, r \neq 1} p \sum_{r \in \mathbb{Q}_1, r \neq 1} 0 = 0$ 이므로 모순이다.
2. 마찬가지로  $0 < p \leq 1$ 일 때에는  $\sum_{r \in \mathbb{Q}_1, r \neq 1} p = \infty$ 이므로 모순이다.

즉  $H$ 는 사건이 아닌 셈이 되고  $P\{H\}$ 가 존재하지 않는다. 이상의 결과를 다음 정리로 요약해본다.

**Theorem 2.5.1** (연속 표본공간에서 시그마-체로 멱집합을 쓰지 않는 이유). 셀 수 없는 표본공간  $[0, 1]$ 에서 시그마-체  $2^{[0, 1]}$ 을 고려할 경우  $P\{[a, b]\} = b - a$ , for all  $0 \leq a \leq b \leq 1$ 과  $P\{A \oplus r\} = P\{A\}$ , for all  $A \subseteq [0, 1]$  and  $0 < r < 1$ 을 동시에 만족하는  $P : 2^{[0, 1]} \rightarrow [0, 1]$ 은 존재하지 않는다.  $\square$

다시말하면, 모든  $A \subseteq [0, 1]$ 에 대해 균등확률  $P\{A\}$ 를 정의할 수 없다는 것이다.

### 2.5.2 실수공간에서 보렐 시그마-체 (Borel sigma-field on $\mathbb{R}$ )

따라서, 모든 interval을 포함하는 시그마-체들 중 가장 작은 시그마-체  $\mathcal{F}$ 를 찾는 것이 이상적일 것이다. 즉 우리는  $\sigma(\text{intervals})$ 를 찾고자 하는 것이다.

여기서 잠시  $\mathbb{R}$ 에서의 보렐 시그마-체 (Borel sigma-algebra)에 대해 살펴보자.  $\mathbb{R}$ 에서의 모든 열린 집합(open set)들의 모임을  $\mathcal{O}$ 라고 하자. 그러면  $\mathcal{O}$ 는 시그마-체가 아니다. (왜냐하면  $A \in \mathcal{O}$ 이면  $A^C$ 는 닫힌 집합이고 따라서  $A^C \notin \mathcal{O}$ 이다.)

**Definition 2.5.2** (보렐 시그마-체).  $\mathbb{R}$ 에서의 보렐 시그마-체 (Borel sigma-field, Borel sigma-algebra)  $\mathcal{B}$ 는  $\mathcal{B} = \sigma(\mathcal{O})$ 로 정의한다.  $\square$

결론은  $\mathbb{R}$ 에서의 보렐 시그마-체를 interval을 포함하는 시그마-체로 만들 수 있다는 것이다. 그 전에 증명을 위해 한 가지 정리를 언급하겠다.

**Theorem 2.5.3** (열린 집합과 열린 구간들).  $E \subseteq \mathbb{R}$ 이 열린 집합이라고 하자. 그러면 기껏해야 셀 수 있는 정도로만 많은 (at most countably many) 열린 구간들 (open intervals)

$I_j, j = 1, 2, \dots$ ,가 존재해 다음을 만족한다.

$$E = \bigcup_{j=1}^{\infty} I_j.$$

□

증명. 이 정리의 증명은  $E = \bigcup_{j=1}^{\infty} I_j$ 를 만족하는  $I_j$ 들이 있음을 보이고 이것이 (1) at most countably many (2) disjoint (3) open (4) intervals 임을 보이면 된다. 증명 방법은 동치 관계를 이용하는 것이다.  $a, b \in E, (a < b)$  일 때 열린구간  $(a, b) \subseteq E$ 이면  $a \sim b$ 로 놓는다. 그러면  $E$ 의 disjoint union of classes는 equivalence relationship partitions  $E$ 가 된다. 아직 이것들이 countably many한지 모르므로 이들을  $I_j, j \in J, J$  is an arbitrary index set 이라고 하자. 임의의  $a_j < b_j \in I_j$ 에 대해  $a_j \sim b_j$ 이고  $(a_j, b_j) \in I_j$ 이므로  $I_j$ 는 interval 이다. 또  $x \in I_j$ 를 임의로 뽑았을 때  $x \in E$ 이고  $E$ 가 open이므로,  $(x - \epsilon, x + \epsilon) \subseteq E$ 를 만족하는  $\epsilon > 0$ 이 모든  $x$ 에 대해 존재함을 알 수 있고 ahems  $a \in (x - \epsilon, x + \epsilon)$ 에 대해  $a \sim x$ 이므로  $x$ 의  $\epsilon$ -근방은  $I_j$ 에 포함됨을 알 수 있어  $I_j$ 는 open이다. 마지막으로 모든  $I_j$ 는 적어도 하나의 유리수를 포함하고 있어 이를 통해  $I_j$ 의 갯수는 countably many임을 알 수 있다. □

**Theorem 2.5.4** (실수 구간에서의 보렐 시그마-체의 생성).  $\mathbb{R}$ 에서의 보렐 시그마-체 (*Borel sigma-field, Borel sigma-algebra*)  $\mathcal{B}$ 는  $(-\infty, a], a \in \mathbb{Q}$ 로 생성할 수 있다. □

증명.  $\mathcal{O}_0$ 을  $\mathbb{R}$ 에서의 모든 열린 구간(open interval)들의 collection이라고 하자. 앞선 정리에 의해  $\mathbb{R}$ 에서의 모든 열린 집합은 기껏해야 셀 수 있을 정도의 열린 구간의 합집합으로 나타낼 수 있으므로  $\sigma(\mathcal{O}_0) = \mathcal{B}$ 가 된다.  $\mathcal{D}$ 를  $(-\infty, a], a \in \mathbb{Q}$ 인 구간들의 collection이라고 하자. 그리고 구간  $(a, b)$ 를  $(a, b) \in \mathcal{O}_0, a, b \in \mathbb{Q}, a < b$ 와 같이 정의하자. 여기서

$$a_n = a + \frac{1}{n} \text{ and } b_n = b - \frac{1}{n}$$

으로 놓으면

$$(a, b) = \bigcup_{n=1}^{\infty} (a_n, b_n] = \bigcup_{n=1}^{\infty} \{(-\infty, b_n] \cap (-\infty, a_n]^c\}$$

이다. 이것은  $(a, b) \in \sigma(\mathcal{D})$ 임을 의미한다. 즉  $\mathcal{O}_0 \subseteq \sigma(\mathcal{D})$ 이고  $\sigma(\mathcal{O}_0) \subseteq \sigma(\mathcal{D})$ 이다. 그러나  $\mathcal{D}$ 의 모든 원소는 닫힌 집합이고 이는  $\sigma(\mathcal{D}) \subseteq \mathcal{B}$ 를 의미한다. 따라서

$$\mathcal{B} = \sigma(\mathcal{O}_0) \subseteq \sigma(\mathcal{D}) \subseteq \mathcal{B}$$

이므로  $\sigma(\mathcal{D}) = \mathcal{B}$ 이다. □

이제  $[0, 1] \subset \mathbb{R}$ 에서의 보렐 시그마-체  $\mathcal{B}_1$ 을 생각해보자. 이는  $\mathcal{B}$ 와 마찬가지로  $[0, 1]$ 의 열린 부분집합들의 collection으로부터 생성된 시그마-체라고 생각할 수 있다. 한 가지 주의해야 할 점은  $\mathcal{B}_1$ 은  $\mathcal{B}$ 의 부분 시그마-체는 아니라는 것이다. 시그마-체에서 포함 관계를 얘기하려면 두 시그마-체의 표본공간이 같아야 한다.

### 2.5.3 실수공간에서 확률측도의 구성 (construction of a probability measure on $\mathbb{R}$ )

이제 앞서 언급한 보렐 시그마-체  $\mathcal{B}_1$ 을 이용해 균등한 확률을 만드는 작업을 진행할 것이다. 전략은 먼저 **단조축정리 (monotone class theorem)**을 증명한 후 그것의 따름정리를 이용해 모든  $0 \leq a \leq b \leq 1$ 에서 균등확률  $P\{[a, b]\} = b - a$ 를 정의한 후 이것을  $[0, 1]$ 의 보렐집합으로 확장하는 것이다.

$\mathcal{F}_0$ 를 표본공간  $\Omega$ 에서의 부분집합들의 collection이라고 하자. 그러면  $\mathcal{F}_0$ 에서 확률측도  $P : \mathcal{F}_0 \rightarrow [0, 1]$ 을 정의한다. 그리고  $\mathcal{F} = \sigma(\mathcal{F}_0)$ 의 확률측도  $Q : \mathcal{F} \rightarrow [0, 1]$ 가  $Q\{A\} = P\{A\} \forall A \in \mathcal{F}_0$ 인 성질을 가진체로 정의된다는 것을 보인다. 마지막으로 이  $Q$ 가 유일함을 보일 것이다.

그 전에 몇 가지 개념들에 대해 정의하겠다.

**Definition 2.5.5** (유한한 교집합에 대해 닫혀있다). 표본공간  $\Omega$ 의 부분집합들의 class  $\mathcal{C}$ 가

$$\bigcap_{i=1}^n A_i \in \mathcal{C} \quad \text{for every } n < \infty \text{ and } A_1, \dots, A_n \in \mathcal{C}$$

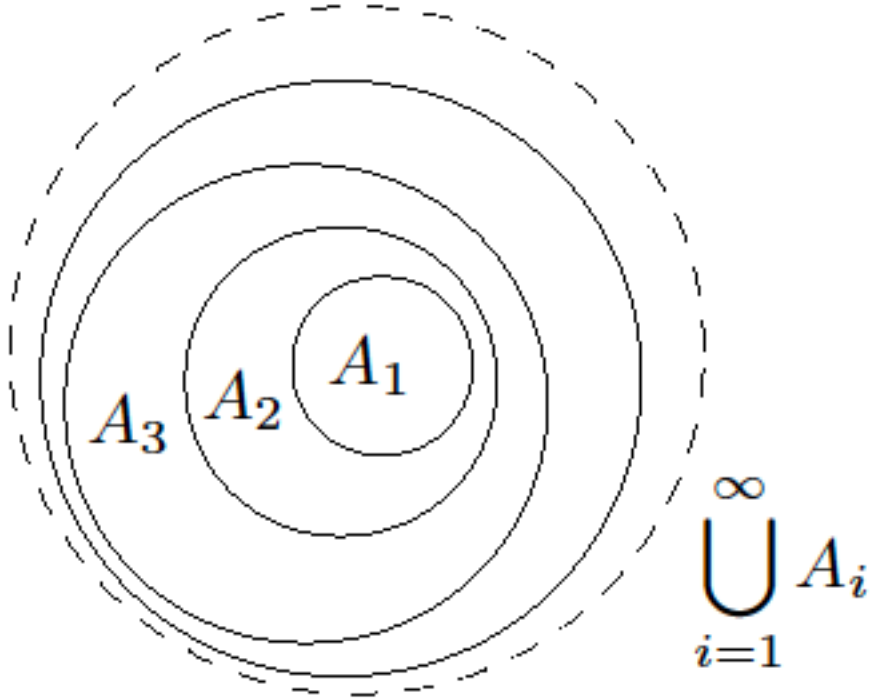
를 만족하면 **유한한 교집합에 대해 닫혀있다 (closed under finite intersection)**라고 한다. □

**Definition 2.5.6** (증가하는 극한에 대해 닫혀있다). 표본공간  $\Omega$ 의 부분집합들의 class  $\mathcal{C}$ 가

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{C} \quad \text{for every collection } A_1, A_2, \dots \in \mathcal{C} \text{ with } A_1 \subseteq A_2 \subseteq \dots$$

를 만족하면 **증가하는 극한에 대해 닫혀있다 (closed under increasing limits)**라고 한다. □

**Definition 2.5.7** (유한한 차에 대해 닫혀있다). 표본공간  $\Omega$ 의 부분집합들의 class  $\mathcal{C}$ 가 모든  $A, B \in \mathcal{C}, A \subseteq B$ 에 대해  $B \setminus A \in \mathcal{C}$ 를 만족하면 **유한한 차에 대해 닫혀있다 (closed under finite differences)**라고 한다. □



**그림 2.3:** Closed under increasing limits.

다음은 단조족에 대한 정의다.

**Definition 2.5.8** (단조족). 표본공간  $\Omega$ 의 부분집합들의 class  $\mathcal{C}$ 가 closed under countable increasing set이거나 closed under countable decreasing set이면, 즉  $E_1, E_2, \dots \in \mathcal{C}$ 일 때

$$E_i \uparrow E \text{ or } E_i \downarrow E \implies E \in \mathcal{C}$$

이면  $\mathcal{C}$ 를 **단조족 (monotone class)**이라고 한다. □

이에 대한 증명은 (?)의 Appendix 부분을 참고하자.



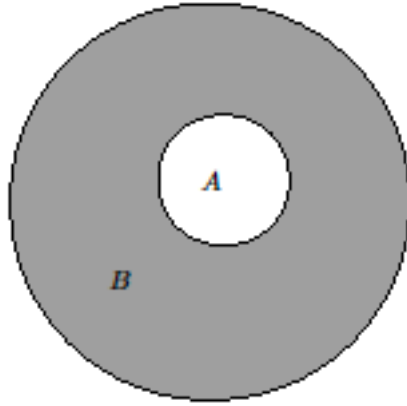


그림 2.4: Closed under finite differences.

## 2.6 측도 (measure)

측도 (measure)란 수학에서, 양 (quantity)이라 개념을 반영하기 위해 만들어진 장치다.  $X$ 를 어떤 집합이라고 하고, 이것의 부분집합  $U \subset X$ 가 있을 때,

### 2.6.1 확률측도 (probability measure)

확률측도는 측도들 중  $P(\Omega) = 1$ 인 측도  $P$ 를 일컫는다.

**Definition 2.6.1** (르베그 측도). 시그마-장  $\mathcal{L}$ 이 있을 때  $C$ 가 어떤 구간(interval)일 경우  $\mu_L(C) = \text{length}C$ 로 정의한 측도  $\mu_L$ 을 르베그 측도 (Lebesgue measure)라고 부른다. 더불어  $\mathcal{L}$ 을 르베그 집합 (Lebesgue set)이라고 부른다. 만약 우리가 르베그 집합을  $[0, 1]$ 로 한정할 경우,  $\mu_L$ 은 확률측도가 된다.  $\square$

## 2.7 라돈-니코딤 정리 (Radon-Nykodim theorem)



### 3 확률변수

**Definition 3.0.1** (가측).  $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$ 가 두 가측공간(measurable)이라고 할 때, 함수  $X : \Omega_1 \rightarrow \Omega_2$ 가 임의의 집합  $E \in \mathcal{F}_2$ 에 대해 집합  $X^{-1}(E)$ 이

$$X^{-1}(E) = \{\omega \in \Omega_1 : X(\omega) \in E\} \in \mathcal{F}_1$$

(또는 위 조건을  $X^{-1}(\mathcal{F}_2) \in \mathcal{F}_1$ ) 일 때  $X$ 가 **가측(measurable)**이라고 한다. (또는  $\mathcal{F}_1$ -가측이라고도 부른다.)  $\square$

**Definition 3.0.2** (확률변수).  $(\Omega, \mathcal{F}, P)$ 가 확률공간(probability space)이라고 하자. 그러면 어떤 실변수 함수

$$X : \Omega \rightarrow \mathbb{R}$$

가  $(\Omega, \mathcal{F})$ 에서  $(\mathbb{R}, \mathcal{B})$ 로 가는 가측함수(measurable function)일 때(여기서  $\mathcal{B}$ 는  $\mathbb{R}$ 에서의 보렐 시그마-체) 이  $X$ 를 **확률변수(random variable)**라고 부른다.  $\square$

(확률변수에 대한 설명 보충 필요)

#### 3.1 보렐-칸텔리 따름정리

**Lemma 3.1.1** (보렐-칸텔리 따름정리). (*Borel-Cantelli lemma*) 만약  $\sum_{n=1}^{\infty} P(A_n) < \infty$ 라면

$$P(A_n i.o.) = 0$$

이다.  $\square$

**Lemma 3.1.2** (제 2 보렐-칸텔리 따름정리). (*Second Borel-Cantelli lemma*) 만약 사건  $A_n$ 들이 독립이라면  $\sum P(A_n) = \infty$ 는  $P(A_n i.o.) = 1$ 임을 내포한다.  $\square$



## 4 확률변수의 수렴

(?)의 내용을 따라간다.

### 4.1 거의 확실한 수렴 (Almost sure convergence)

$X_1, X_2, \dots$ 가 확률공간  $(\Omega, \mathcal{F}, P)$ 에서의 확률변수의 수열이라고 하자. 고정된  $\omega$ 에 대해  $X_n(\omega) = x_n, n = 1, 2, \dots$ 은 숫자의 수열이라고 하자. 각  $\omega$ 에 대해  $X_n(\omega)$ 가 수렴할 수 있지만 극한  $X(\omega)$ 는  $\omega$ 에 따라 다를 수 있다. 예를 들면,  $(\Omega, \mathcal{F}, P) = ([0, 1], \mathcal{B}_{[0,1]}, \mu_L)$  이고

$$X_n(\omega) = \omega^n \quad (4.1.1)$$

이다. 그러면  $n \rightarrow \infty$ 일 때  $X_n(\omega) \rightarrow I(\omega = 1)$ 이다. 그런데 어떤  $\omega$ 에 대해서는  $X_n(\omega)$ 는 극한이 없거나 무한대의 극한을 갖을 수 있다. 예를 들면 앞선 식 (4.1.1)을 다음과 같이 바꾸는 것이다.

$$X_n(\omega) = (-\omega)^n \quad (4.1.2)$$

그러면  $\omega < 1$ 일 때  $n \rightarrow \infty$ 이나  $\omega = 1$ 일 때는 극한이 존재하지 않는다.

식 (4.1.1)과 (4.1.2)에서의 행동이 다르다고 하더라도  $\{\omega = 1\}$ 이 확률 0을 갖는다면 다른 행동을 무시할 수 있을 것이다. 이것을 확장시키면 확률 0인 집합들을 무시하는 것으로 이해할 수 있고, 거의 확실한 수렴 (Almost sure convergence)의 정의를 이끈다.

## 4.2 확률수렴 (Convergence in probability)

## 4.3 $L_p$ 수렴 (Convergence in $L_p$ )

## 4.4 분포수렴 (Convergence in distribution)

**분포수렴 (Convergence in distribution)**은 분포함수의 관계를 다룬다는 것에서 확률 변수들의 관계를 고려하는 앞 수렴들과는 다른 타입의 수렴이라고 할 수 있다.  $X_n$ 의 분포함수  $F_n(x)$ 가  $X$ 의 분포함수  $F(x)$ 로 수렴할 때 우리는  $X_n$ 이  $X$ 에 대해 분포적으로 가까워진다고 말한다. 가장 단순한 예로  $X_n = 1 + \frac{1}{n}$ 이고  $X$ 는 1인 경우를 생각해 볼 수 있다. 이 때  $X_n$ 의 분포함수  $F_n(x)$ 는 다음과 같은 분포함수  $F(x)$ 로 가까워지는 것처럼 보인다.

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 & \text{if } x \geq 1 \end{cases} \quad (4.4.1)$$

그러나  $F_n(x) = P(X_n \leq x)$ 는

$$F_n(x) = \begin{cases} 0 & \text{if } x < 1 + \frac{1}{n} \\ 1 & \text{if } x \geq 1 + \frac{1}{n} \end{cases} \quad (4.4.2)$$

이며 이는

$$\begin{cases} 0 & \text{if } x \leq 1 \\ 1 & \text{if } x > 1 \end{cases} \quad (4.4.3)$$

로 수렴한다. 식 (4.4.1)과 식 (4.4.2)의 분포함수는  $x = 1$ 에서 일치하지 않는다. 따라서,  $F_n(x)$ 가 모든  $x$ 에서  $F(x)$ 로 수렴하는 것은 너무 강한 조건으로 보인다. 분포수렴을 정의할 때에는 앞선 예의  $x = 1$ 처럼 불연속인 점들을 제외한 연속인 점들  $x$ 에서  $F_n(x) \rightarrow F(x)$ 가 되는 것으로 정의한다.