통계공부와 관련된 글들

Seoncheol Park

2017-08-31

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 A가 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. □

1.2 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 1.2.1 (점별수렴)**.** $\{f_n(x)\}_{n=1}^{\infty}$를 실함수의 수열이라고 하자. 그러면 모든 $x \in \mathbb{R}$에 대해

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

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

$$f_n \stackrel{pw}{\to} f$$

로 표현하기로 한다. □

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

Definition 1.2.2 (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_ϵ 이 존재할 때 수열 $\{f_n(x)\}_{n=1}^{\infty}$ 이 f 에 **균등수렴 (uniform convergence)** 한다고 한다. 여기서는 $n \to \infty$ 일 때

$$f_n \stackrel{u}{\to} f$$

로 표현하기로 한다. □

(두 수열의 차이점 설명)

1.3 연산자들과 노름 (operators and norms)

1.3.1 직합 (direct sum)

Definition 1.3.1 (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 1.3.1 (직합의 예)**.**

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

□

1.3.2 크로네커 곱 (Kronecker product)

Definition 1.3.2 (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 1.3.2 (크로네커 곱의 예)**.** 다음은 크로네커 곱의 한 예이다.

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

□

1.3.3 텐서곱 (tensor product)

1.3.4 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.

1.3.5 노름 (norm)

1.3.5.1 벡터 노름 (vector norm)

Definition 1.3.3 (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 α 와 임의의 $\mathbf{x} \in X$ 에 대해
$$\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$$
가 성립한다.

□

1.3.5.2 행렬 노름 (matrix norm)

1.3.6 거리 (metric)

Definition 1.3.4 (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 필요충분조건은 $x = y$ 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 이니다.

1.4 행렬의 분해 (matrix decomposition)

1.4.1 고유값 분해 (eigenvalue decomposition)

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

1.4.2 스펙트럼 분해 (spectral decomposition)

$p \times p$ 대칭행렬 A에 대한 **스펙트럼 분해 (spectral decomposition)**는 다음과 같다. $p \times p$ 대칭행렬 A는 직교행렬 P에 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 P는 $P = [e_1, \dots, e_p]$로 이루어지며, Λ는 A의 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 Λ는 $P^T A P = \Lambda$이다.

1.4.3 특이값분해 (SVD)

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

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

이 때

- U: A의 left singular vector로 이루어진 $m \times m$ 직교행렬 (orthogonal matrix)
- Σ: 주 대각성분이 $\sqrt{\lambda_i}$로 이루어진 $m \times n$ 직사각 대각행렬 (diagonal matrix)

- V: A의 right singular vector로 이루어진 $n \times n$ 직교행렬(orthogonal matrix)

행렬 A의 계수(rank)가 k라고 할 때,

- $U = [u_1, \dots, u_k, \dots u_m]$는 AA^T를 고유값분해(eigenvalue decomposition)로 직교대각화하여 얻은 $m \times m$ 직교행렬(orthogonal matrix)이며, 특히 $[u_1, \dots, u_k]$를 **좌특이벡터(left signular 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 signular vector)**라고 한다.

- Σ는 $A^T A$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 A를 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, V가 직교행렬이면 $UU^T = I$, $VV^T = I$가 성립한다.

1.4.4 특이값분해와 고유값분해의 관계

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

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

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

1.4.5 특이값 분해의 기하학적 의미

위키피디아[1]를 참고하자.

1.5 기저 (basis)

1.5.1 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 consts m

[1] https://en.wikipedia.org/wiki/Singular_value_decomposition

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

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

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

(동익이형 박사논문 47쪽)

1.5.2 Radial basis function

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

- Gaussian

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

- Multiquadric

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

- Inverse quadratic

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

- Inverse multiquadric

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

1.6 공간 (space)

1.6.1 L2 공간 (L2 space)

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

1.6.2 바나흐공간 (Banach space)

1.6.3 힐버트공간 (Hilbert space)

1.6.4 Sobolev 공간 (Sobolev space)

1.6.5 Besov 공간 (Besov space)

1.6.6 Reproducing kernel Hilbert space

1.7 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_i는 i번째 변수에 대한 표준편차
- 마할라노비스 거리 (Mahalanobis distance): $d(x,y) = (x-y)^T\mathbf{\Sigma}^{-1}(x-y)$, 여기서 Σ는 공분산행렬
- 체비셰프 거리 (Chebychev distance): $d(x,y) = \max_{i=1,\ldots,p}|x_i - y_i|$

다음 거리들은 유클리드 거리와 더불어 공간통계에서 많이 쓰이는 것들이다.

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

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)** Ω는 그 실험에서 나올 수 있는 모든 결과들의 집합이다. □

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. □

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

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

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

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

로 볼 수 있다. 이러한 표본공간을 **(무한) 이산 표본공간 (infinite discrete sample space)**라고 한다. □

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

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

이다. 이러한 표본 공간은 **연속 표본공간 (continuous sample space)**라고 한다. □

Definition 2.1.2 (사건)**. 사건 (event)**란 표본공간 Ω 의 임의의 부분집합 (subset) 을 의미한다. □

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.

□

2.2 시그마-체 (sigma-field)

앞서 표본공간 Ω 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) Ω 의 non-empty collection (즉 Ω 의 subset 들의 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. □

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

Definition 2.2.2 (시그마-체)**.** 어떤 집합 (set) Ω의 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. $A_1, A_2, \ldots \in \mathcal{F}$ 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. □

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,

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

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

3. $\mathcal{F}$가 *non-void*일 경우에는 모든 체 또는 시그마-체가 A를 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.

□

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

- Ω의 멱집합 (power set, 어떤 집합의 모든 부분집합을 모은 집합) 또한 시그마-체이며 이는 Ω의 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.

□

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

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

와 $(a, b]$ 의 모든 finite union을 포함한다고 하자. 그리고 $[z]$를 z와 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

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

이다. 따라서 $\mathcal{F}$는 시그마-체가 아니다. □

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

2.3 생성기들 (generators)

시그마-체에 대해 좀 더 자세히 살펴보기 위해, **생성기 (generator)**에 대해 알아보자. 표본공간 Ω의 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

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

인 $\mathcal{F}$가 존재하고 이를 **$\mathcal{A}$로부터 생성된 시그마-체 (sigma alegbra genearted 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.

□

2.4 확률공간(probability space)

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

Definition 2.4.2 (확률측도)**.** 가측공간 $(\Omega, \mathcal{A})$가 주어졌을 때 **확률측도(probability measure)** P는 $P : \mathcal{A} \to [0, 1]$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 disjoint라고 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. □

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 Ω가 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 E를 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 Ω의 모든 interval들을 포함하는 시그마-체를 만들고 싶어한다. 예를 들면, $\Omega = [0, 1]$일 때

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

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

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 probablity of power set on continous sample space)

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

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

을 만족하길 원한다. 또한 특별히

$$P\{a\} = 0, \qquad \text{for every } 0 \le a \le 1$$

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

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

를 만족해야 한다.

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

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

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

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

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

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

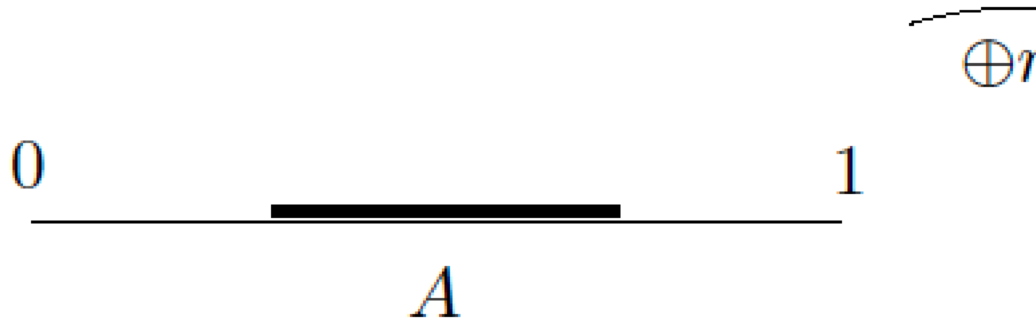

그림 2.1: Shift invariance.

로 정의할 수 있다.

“wrapping around”를 이용해 A를 r만큼 이동해도 길이가 보존되기 때문에, 확률 또한

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

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. x와 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.

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

H를 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

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

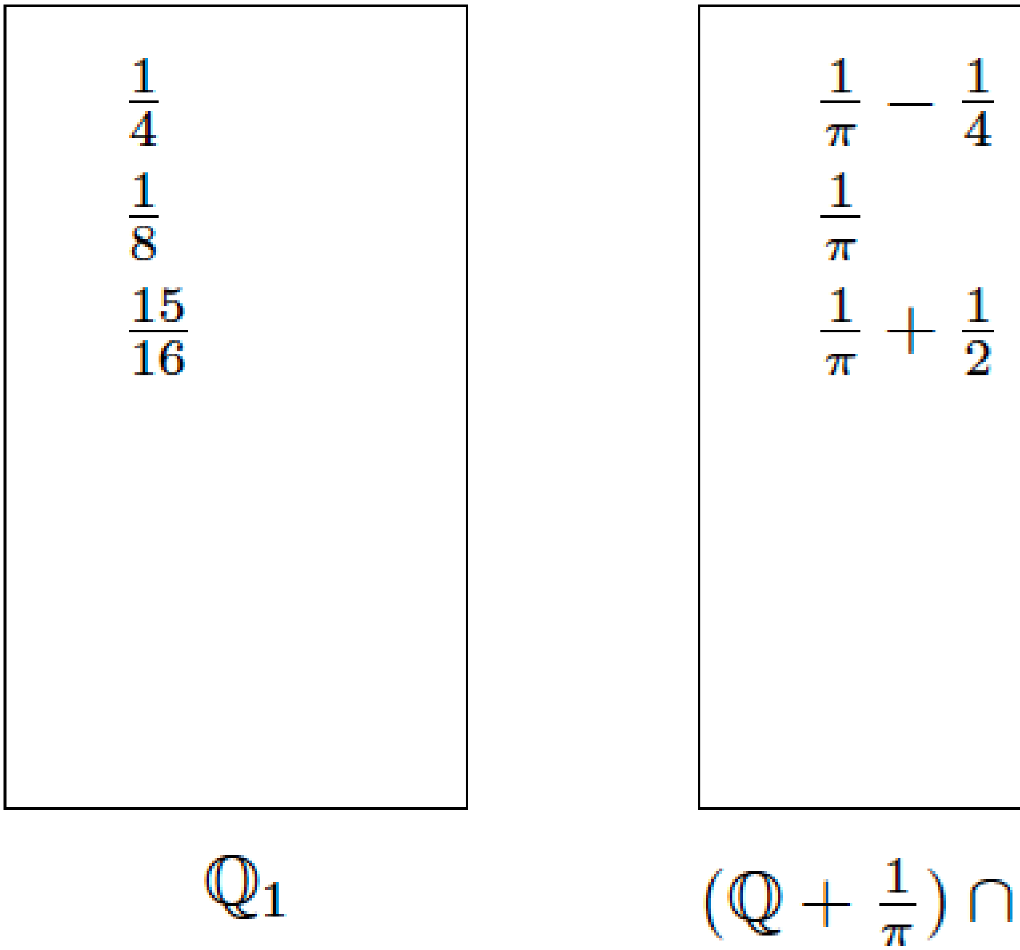

그림 2.2: Collection of disjoint unions on [0,1]

로 표현할 수 있다. 그러면

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

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 $P\{[a,b]\} = b - a,\ \textit{for all}\ 0 \leq a \leq b \leq 1$ 과 $P\{A\oplus r\} = P\{A\},\ \textit{for all}\ A \subseteq [0,1]\ \textit{and}\ 0 < r < 1$ 을 동시에 만족하는 $P : 2^{[0,1]} \to [0,1]$ 은 존재하지 않는다. □

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 R)

따라서, 모든 interval을 포함하는 시그마-체들 중 가장 **작은** 시그마-체 $\mathcal{F}$ 를 찾는 것이 이상적일 것이다. 즉 우리는 σ(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. □

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)*