1 Measure Theory

● 여기서는 듀렛책 챕터 1.1 ∼ 1.3 사이의 내용을 다룬다. 이 내용들은 probability spaces, distributions, random variable에 대한 내용인데 전반적으로 적분이전의 measure theory에 관한 내용이다. 개인적으로 책이 상당히 두서없이 정리되어 있다고 생각한다. 그래서 나도 두서없이 내맘대로 정리하였다.

● 확률변수 X 에 대한 정의를 생각하여 보자. 확률변수는 본질적으로 메저러블-맵핑이므로 메저러블-매핑의 정의에 대하여 알아보자. 어떠한 맵핑 X 가 메저러블-스페이스 $(\Omega, \mathcal{F})$ 와 메저러블-스페이스 $(S, \mathcal{S})$ 를 이어주는 메저러블-맵핑이라는 의미는 1) X 가 Ω 에서 S 로 가는 함수이고 즉 $X : \Omega \to S$ 이고 2) X^{-1} 가

$$\forall B \in \mathcal{S} : X^{-1}(B) \in \mathcal{F}$$

를 만족한다는 의미이다.

● 확률변수 X 라는 것을 정의하기 위해서는 반드시 두 메저러블-스페이스가 필요하다. 확률변수 X 단독으로 정의될 수 없다. 이것은 비유하면 "남친"의 정의와 비슷하다. 남친은 혼자 될 수 있는 것이 아니다. 즉 남친은 단독으로 정의될 수 없고 "XXX의 남친"과 같은 식으로만 정의된다. 예를 들면

"저는 남친이에요"

라고 말하는것은 올바른 정의가 아니며

"저는 XXX의 남친이에요"

라고 말하는 것이 올바른 정의이다. 이처럼

"X 는 확률변수에요"

라고 말하는 것은 올바른 정의가 아니며 (하지만 유감스럽게도 대부분 이렇게 말하고, 그래서 확률변수가 무엇인지 헷갈려한다. 그리고 내가 매년 이러고 있다..)

"X 는 두 메저러블-스페이스 $(\Omega, \mathcal{F})$ 와 $(S, \mathcal{S})$ 를 연결하는 확률변수에요"

라고 말하는것이 올바른 정의라는 것이다. 요런 느낌의 정의를 확률론에서는 많이 가지고 있다. 가령 예를들면 시그마필드 $\mathcal{F}$ 는 홀로 정의될 수 없고 전체집합 Ω 가 존재해야지 정의된다. 그리고 임의의 메저러블셋 A 은 시그마필드 $\mathcal{F}$ 가 있어야지 정의될 수 있다. 즉 아래와 같이 말한다.

"$\mathcal{F}$ 는 Ω 에 대한 시그마필드에요"

혹은

"집합 A는 시그마필드 $\mathcal{F}$에 대해서 메저러블해요"

와 같은 식으로 말이다.

● 확률변수는 사실 메저러블-맵핑중에서 매우 특수한 경우이다. 왜냐하면 "이미지(=image)"에 해당하는 메저러블-스페이스 $(S, \mathcal{S})$가 $(\mathbb{R}, \mathcal{R})$로 고정되어있기 때문이다. 여기에서 $\mathcal{R}$은 보렐클라스이다. 그러면 사실 확률변수 X를 정의하는데 Ω, $\mathcal{F}$, S, $\mathcal{S}$를 일일히 정의할 필요는 없어진다. $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{R})$은 이미 정해졌고 Ω는 $\mathcal{F}$를 정의하면 사실 같이 정의된 셈이나 마찬가지다($\mathcal{F}$에서 가장 큰 원소가 Ω일테니깐!). 종합하면 확률변수 X를 정의할때는 $\mathcal{F}$만 정의하면 된다. 따라서

"X는 두 메저러블-스페이스 $(\Omega, \mathcal{F})$와 $(S, \mathcal{S})$를 연결하는 확률변수에요"

라고 말하는것이 정확하지만

"X는 $\mathcal{F}$에서 정의된 확률변수에요"

라고 말해도 틀린말은 아니라는 것이다. 그리고 이렇게 말하는 것을 더 좋아하는 사람들이 많다(간단하니깐). 그리고 이것을 기호로는

$$X \in \mathcal{F}$$

라고 쓴다.

● 아마 나는 이 포스팅을 읽을때마다 보렐클라스가 무엇인지 다시 헷갈리기 시작할것이다(나 자신에 대한 끝없는 불신..). 그래서 보렐클라스에 대해서 다시 정의하고 넘어가겠다. 보렐클라스 $\mathcal{R}$이란 $\mathbb{R}$의 모든 오픈셋을 포함하는 시그마필드라는 것이다. 좀 더 엄밀하게 써보자. 집합 O를 $\mathbb{R}$에 속하는 임의의 오픈셋이라고 하자. 즉

$$O \subset \mathbb{R} \quad and \;\; O \;\; is \;\; open \;\; set$$

이라고 하자. 이러한 집합 O들을 모아놓은 클라스를 $\mathcal{O}$를 정의하자. 즉

$$\mathcal{O} = \{O : O \subset \mathbb{R}, \quad O \;\; is \;\; open \;\; set\}$$

이다. $\mathcal{R}$은 $\mathcal{O}$를 포함하는 가장 작은 시그마필드이다. 즉 (1) $\mathcal{O} \subset \mathcal{R}$이고 (2) "$\mathcal{R}$ is σ-field." 이다. 이것을 기호로는 아래와 같이 쓴다.

$$\mathcal{R} = \sigma(\mathcal{O})$$

그리고 $\mathcal{R}$을 $\mathcal{O}$에 의해서 제너레이트된 시그마필드라고도하며 $\mathcal{O}$가 $\mathcal{R}$을 생성했다고 말하기도 한다.

• 보렐클라스는 우리가 상상할 수 있는 대부분의 "subset of $\mathbb{R}$"을 포함한다. 가령 예를들면 보렐클라스는 점집합, 인터벌, 레이와 같은 집합들을 포함하는데 이는 이러한 집합들이 (a, b)의 카운터블 유니온꼴로 표현될 수 있다는 것을 알면 쉽게 이해할 수 있다. 이러한 집합을 $\mathcal{R}$-메저러블한 집합이라고 표현한다. 간단히 줄여서 보렐셋이라고 표현하기도 한다. 하지만 "subset of $\mathbb{R}$"임에도 불구하고 가끔 가다가 보렐클라스의 원소가 아닌 이상한 집합이 있을 수도 있다. 이런 집합을 보렐클라스의 원소가 아닌 집합 혹은 $\mathcal{R}$-메저러블하지 않은 집합 혹은 보렐셋이 아닌 집합이라고 한다. 대표적으로 비탈리집합이 있다.

• 보렐셋 즉 $\mathcal{R}$-메저러블한 집합은 $\mathbb{R}$에서 **"길이를 잴 수 있는 집합"**이라고 이해해도 괜찮다. 가령 예를 들면 $[0, 1]$의 길이는 1이고 $\emptyset$의 길이는 0이다. 또한 $\{0\}$와 같은 집합의 길이는 0, 유리수들의 집합의 길이는 0, 그리고 $[0, 1]$사이에 포함된 무리수들의 집합을 길이로 재면 1과 같은식으로 정의할 수 있다. 따라서 위에서 언급한 집합들은 모두 잴 수 있는 집합이며, $\mathcal{R}$-메저러블한 집합이다.

• 참고로 위와 같이 길이를 재는 함수를 **르벡메저**라고 한다. 르벡메저는 $\mathcal{R}$에 속한 임의의 집합에 대하여서도 이러식으로 **모순없이** 길이를 정의할 수 있다. 여기에서 모순없이 길이를 잴 수 있다는 말이 의미하는 것은 (1) 공집합의 길이는 0이며 (2) $\mathcal{R}$에 속하는 어떠한 집합의 길이도 양수이어야 하며 3) $\mathcal{F}$에 속하는 두 집합이 서로소라면(즉 $[0, 1]$, $[3, 4]$와 같은 집합) 두 집합의 합친집합의 길이는 각각의 길이의 합과 같아야 한다는 것이다(즉 $[0, 1] \cup [3, 4]$의 길이는 $[0, 1]$의 길이와 $[3, 4]$의 길이의 합과 같아야한다는것임). 어떠한 클라스 $\mathcal{F}$에서 이러한 위와 같은 조건을 만족하는 함수 μ가 있다고 하면 이때 μ는 $\mathcal{F}$의 모든 원소들의 크기(혹은 길이)를 **모순없이** 정의할 수 있다. 이러한 함수 $\mu : \mathcal{F} \to [0, \infty]$를 $\mathcal{F}$에서의 메저라고 한다. 참고로 **르벡메져** λ는 아래와 같이 정의되는 함수이다.

$$\lambda((a, b]) = b - a$$

이러한 함수 λ는 보렐클래스 $\mathcal{R}$에 있는 모든집합에 대하여 **모순없이** 길이를 정의할 수 있다. 따라서 λ는 $\mathcal{R}$에 대한 메져이다.

• 다시 X로 관심을 돌려보자. X는 기본적으로 Ω와 $\mathbb{R}$을 연결하는 맵핑이다. 이것이 $(\Omega, \mathcal{F})$와 $(\mathbb{R}, \mathcal{R})$을 연결하는 맵핑, 즉 메저러블 맵핑이 되려면 어떻게 되어야 할까? 해답은 매우 간단하다. **Ω의 원소들을 근원사건으로 해석하여 맵핑하지말고 $\mathcal{F}$의 (공집합을 제외한) 가장 작은 원소들을 근원사건으로 해석하여 맵핑시키면 된다.** 이제 이게 무슨말인지 찬찬히 살펴보자. 예를들어

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

이라고 하자. 즉 주사위를 던져서 나오는 눈들의 집합이 Ω이다. Ω의 원소를 기준으로 이해한 근원사건은 다음과 같다.

- 주사위를 던져서 눈이 1이 나올 경우
- $\dots$
- 주사위를 던져서 눈이 6이 나올 경우

이제 Ω에 대한 시그마필드를 아래와 같이 정의하자.

$$\mathcal{F} = \{\emptyset, \{1,3,5\}, \{2,4,6\}, \Omega\}.$$

즉 $\mathcal{F}$의 원소는 (1) 공집합 (2) 전체집합 그리고 (3) 주사위를 던져서 짝수가 나오는 경우를 모은 집합 (4) 주사위를 던져서 홀수가 나오는 경우를 모은 집합이다. 이 경우 $\mathcal{F}$의 (공집합을 제외한) 가장 작은 원소를 기준으로 이해한 근원사건은 다음과 같다.

- 주사위를 던져서 짝수가 나올 경우
- 주사위를 던져서 홀수가 나올 경우

X가 단지 Ω에서 $\mathbb{R}$의 맵핑이 아니라 $(\Omega, \mathcal{F})$에서 $(\mathbb{R}, \mathcal{R})$로 메저러블맵핑이 되려면 **$\mathcal{F}$의 (공집합을 제외한) 가장 작은 원소를 기준으로 맵핑을 해야한다.** 즉

$$X(\{1\}) = 1$$

$$\dots$$

$$X(\{6\}) = 6$$

와 같이 맵핑을 하면 안되고

$$X(\{1\}) = X(\{3\}) = X(\{5\}) = 1$$

$$X(\{2\}) = X(\{4\}) = X(\{6\}) = 0$$

와 같은 식으로 맵핑을 해야한다는 것이다. 참고로 전자의 맵핑은

"X: 주사위를 던져서 나오는 눈의 수"

와 같은 식으로 확률변수를 구성한 것이고 후자의 경우는

"X: 주사위를 던져 홀수면 1 짝수면 0"

과 같은 방식으로 확률변수를 구성한 것이다. 전자의 근원사건은 주사위를 던져서 나온 눈의 수이며 후자의 근원사건은 주사위를 던져서 짝수가 나왔는지 홀수가 나왔는지다. 참고로 후자의 경우처럼 X를 맵핑해야 X가 $\mathcal{F}$에서 정의된 랜덤변수 즉 $X \in \mathcal{F}$라고 할 수 있는 것이다.

- $X \in \mathcal{F}$**가 되도록** X**를 정의하기 위해서는** 위처럼 $\mathcal{F}$가 근원사건을 이해하는 방식으로 맵핑 X를 정의해도 되지만 $\mathcal{F}$**보다 작은 어떤 시그마필드** $\mathcal{F}^*$**의 가장 작은 원소를 기준으로 맵핑** X**를 정의해도 된다.** 예를들면 위의 예제처럼

$$\mathcal{F} = \{\emptyset, \{1,3,5\}, \{2,4,6\}, \Omega\}.$$

라고 하고 우리는 $X \in \mathcal{F}$가 되도록 맵핑 X를 잘 정의하고 싶다고 하자. X를 위의 예제처럼 $\mathcal{F}$의 최소원소를 기준으로 맵핑하 아래와 같다.

$$X(\{1\}) = X(\{3\}) = X(\{5\}) = 1$$

$$X(\{2\}) = X(\{4\}) = X(\{6\}) = 0$$

당연히 $X \in \mathcal{F}$은 성립한다. 그렇다면 $\mathcal{F}$보다 작은 집합 $\mathcal{F}^*$, 예를들면

$$\mathcal{F}^* = \{\emptyset, \Omega\}$$

와 같은 시그마필드의 최소원소를 기준으로 X를 맵핑해도 여전히 $X \in \mathcal{F}$가 성립할까? 그러하다. 예컨데 아래와 같은식으로 X를 맵핑해도된다.

$$X(\{1\}) = X(\{2\}) = X(\{3\}) = X(\{4\}) = X(\{5\}) = X(\{6\}) = 1.$$

이렇게 정의하여도 $X \in \mathcal{F}$가 성립한다. 단지 이러한 경우 $\mathcal{F}$는 X를 정의하기 위해서 필요한 최소한의(=가장작은) 시그마필드는 아니게 된다. 이 경우 X를 정의하는데 필요한 최소한의 시그마필드는 $\mathcal{F}^*$가 된다. 이처럼 맵핑 X를 $\mathcal{F}$ 메저러블하도록 만드는데 필요한 최소한의 시그마필드를 기호로 $\sigma(X)$라고 한다. 그리고 이 예제에서는 $\sigma(X) = \mathcal{F}^*$가 된다.

- 이제 $\mathcal{F}$에서 정의된 어떠한 확률변수 X가 $\mathcal{F}$의 모든 사건들을 **잘** 연결하는 어떠한 맵핑이라는건 알겠다. 그런데 확률변수 X가 어떤 **랜덤**한 출력을 주는 함수라는 느낌은 없다. 왜냐하면 말그대로 확률변수 X는 두 메저러블 스페이스 $(\Omega, \mathcal{F})$, $(\mathbb{R}, \mathcal{R})$을 잘 연결하는 어떠한 맵핑일 뿐이기 때문이다. 즉 X는 단순히 함수(혹은 맵핑)이기 때문에 $\omega \in \Omega$가 고정되면 $X(\omega)$의 값도 고정된다. 따라서 X가 어떤 **랜덤**한 출력을 가지도록 하기 위해서는 ω를 랜덤하게 선택하는 수밖에 없다. 이렇게 ω를 랜덤하게 선택할 수 있게 만들어주는 장치가 바로 확률척도 $P : \mathcal{F} \to \mathbb{R}$이다. $P(\{\omega\})$는 Ω에서 $\{\omega\}$가 선택될 확률을 의미한다. 따라서 P는 ω를 **랜덤**하게 선택할수 있게 해주고 그 결과 $X(\omega)$의 출력 역시 **랜덤**하게 나올 수 있도록 해준다. 따라서 우리가 일반적으로 생각하는 확률변수 X를 정의하기 위해서는 $(\Omega, \mathcal{F})$, $(\mathbb{R}, \mathcal{R})$과 더불어서 P가 추가적으로 필요하다. 즉 확률변수 X를 정의하기 위해서는 $(\Omega, \mathcal{F}, P)$, $(\mathbb{R}, \mathcal{R})$이 필요하다. 여기에서 $(\Omega, \mathcal{F}, P)$를 묶어서 확률공간이라고 한다.

• 엄밀하게 말하면 X를 정의하기 위해서는 $(\Omega, \mathcal{F}, P)$, $(\mathbb{R}, \mathcal{R})$가 모두 필요하지만 앞에서 언급하였듯이 $\mathcal{F}$를 알면 Ω를 아는 셈이고 ($\mathcal{F}$에서 젤 큰 집합이 Ω임) $(\mathbb{R}, \mathcal{R})$는 이미 정해진 것이므로 X를 정의하는데 최소한으로 필요한 정보는 $\mathcal{F}$와 P이다. 그런데 이 사실은 $X \in \mathcal{F}$라는 기호를 이상하게 느껴지게 만들 수 있다. 즉 "X를 **well-define** 하는데 $\mathcal{F}$만 필요하다고 하면 된다고 했는데 사실 그게 아니고 P도 필요한것 아니야?" 라고 생각할 수 있다. 하지만 P는 사실 X가 가지는 **랜덤**성을 주기 위한 장치이고 **따라서 P를 잘 못 정의한다고 하여서 X가 랜덤변수가 아니게 되는 경우는 없다**(P가 probability measure가 되도록만 정의하면 됨)는 사실을 유의하면 X를 **well-define**하기 위해서는 P가 필요하지 않음을 알 수 있다. 즉 아래의 기호는 여전히 유용하다.

$$X \in \mathcal{F}$$

• 그럼 P가 가지는 역할은 무엇일까? P는 X는 가지는 분포를 결정하여 준다. 따라서 P는 맵핑 X가 확률변수인지 아닌지를 결정하지는 않지만 (이것은 $\mathcal{F}$가 결정함) X가 **어떠한** 확률변수인지는 결정한다. 즉 P가 정의되면 부수적으로 X가 가지는 여러가지 성질들을 추가적으로 **assume** 할 수 있다. 예를들어 P는 확률변수의 모멘트가 유한한지 그렇지 않은지를 결정할 수 있으며 확률변수들이 독립인지 아닌지도 결정할 수 있다. (예컨데 $P := P_1 \times P_2$를 확률벡터 (X_1, X_2)의 확률측도라고 하자. 여기에서 P_1, P_2는 각각 X_1, X_2의 probability measure이다. 이때 P를 알고있다면 두 확률변수 X_1, X_2 가 독립인지 독립이 아닌지도 알 수 있다.)

• 확률척도 P는 X의 분포에 대한 정보를 담고 있지만 P의 정의역은 $\mathcal{F}$이다. P의 정의역이 보렐셋 $\mathcal{R}$이 아니기 때문에 종종 P를 쓰기 불편할 때가 있다. 그래서 P와 동등한 역할을 하는 또 다른 함수 $\mu^\star : \mathcal{R} \to [0,1]$를 아래와 같이 만든다.

$$\mu^\star := P \circ X^{-1}.$$

이때 $\tilde{\mu}$를 확률변수 X의 **distribution**이라고 정의한다. 이렇게 정의하면 임의의 집합 $B \in \mathcal{R}$에 대하여 아래식이 성립한다.

$$\mu^\star(B) = P(\{\omega : X(\omega) \in B\}) = P(A)$$

여기에서 $A := X^{-1}(B) := \{\omega : X(\omega) \in B\}$로 정의한다. 참고로 이때 $\mu^\star$는 확률측도가 된다(증명은 알아서 하든가 아니면 그냥 믿든가). 따라서 $(\mathbb{R}, \mathcal{R}, \mu^\star)$는 확률공간이 된다. 여기에서 $\mu^\star$가 X에 의해서 정의되므로 확률공간 $(\mathbb{R}, \mathcal{R}, \mu^\star)$

역시 X 에 의해서 정의되는데 이러한 이유로 **확률공간** $(\mathbb{R}, \mathcal{R}, \mu^\bigstar)$**를 X 에 의해서 유도된 확률공간이라고 표현하기도 한다.**

- 참고로 보통 교재에서는 $\mu^\bigstar$ 대신에 간략하게 μ를 쓴다. 나는 일반적인 메져 μ 와 확률변수 $X \in \mathcal{F}$ 의 distribution 으로서의 (probability) measure 를 구분하기 위해서 $\bigstar$를 붙였다.

- 확률변수 X 에 대한 **distribution** $\mu^\bigstar$ 까지 잘 정의하였다면 확률변수 X 에 대한 **distribution function**을 정의할 수 있다. 확률변수 X 의 **distribution function** $F : \mathbb{R} \to [0, 1]$는 아래와 같이 정의한다.

$$F(x) = \mu^\bigstar((-\infty, x]) = P(\{\omega : X(\omega) \leq x\})$$

$F(x)$는 X 의 **cdf**라고 표현하기도 한다. 참고로 $F(x)$는 (1) $F(-\infty) = 0, F(\infty) = 1$을 만족하고 (2) 비감소이며 (3) 오른쪽 연속인 함수가 된다. 그리고 임의의 함수 $F(x)$가 위의 3가지 조건을 만족하면 $F(x)$는 어떠한 확률변수 X 의 **distribution function**이 된다.

- 여기에서 어떠한 확률변수 X 인지 구체적으로 알아보도록 하자. 결론적으로

$$X(\omega) := \sup \{y : F(y) < \omega\}$$

와 같이 하면 이 확률변수는 $F(x)$를 **distribution function**으로 가지는 하나의 확률변수가 된다.

- 만약에 $F(x)$가 x에 대하여 미분가능하다면

$$f(x) = \frac{d}{dx} F(x)$$

를 만족하는 $f(x)$를 정의할수 있는데 이때 $f(x)$를 X 의 **probability density function**이라고 한다. 그럼 언제 $f(x)$가 존재할까? **라돈-니코딤** 정리는 X 의 **distribution**, $\tilde{\mu} : P \circ X^{-1}$ 가 르벡메져 λ 에 대하여 **absolutely continuous** 하다면 X 의 pdf가 존재함을 시사한다.

- 좀 더 자세히 설명해보자. 내가 알고있기로는 결국 모든 $x \in \mathbb{R}$에 대하여

$$\int_{-\infty}^{x} f(\tau) d\tau = F(x)$$

가 만족하는 $f(x)$가 존재함을 보이면 된다. 참고로 위와 같이 고등학교에서 배운 것같은 기호로 정의된 적분은 **리만적분**이라고 보아도 무방하다. 참고로

$$\lambda\Big(\{x : f(x) \text{ is discontinuous on } x\}\Big) = 0$$

아래가 성립하면 $f(x)$는 리만적분가능에서 **르벡적분가능**으로 바뀌며 따라서

$$\int_{-\infty}^{x} f(\tau)d\tau = \int_{(-\infty,x]} f d\lambda$$

이 된다(김김계 해석개론, p.313, 정리 10.3.4). 여기에서 λ는 르벡메져이다. 따라서 모든 $x \in \mathbb{R}$ 에 대하여 $\int_{-\infty}^{x} f(\tau)d\tau = F(x)$ 가 성립하는 f 가 존재함을 보이기 위해서는 모든 $B \in \mathcal{R}$에 대하여 아래식이 성립하게 하는 함수 f 가 존재함을 보이는 것과 같다.

$$\int_B f d\lambda = \mu^{\bigstar}(B)$$

그런데 $\mu^{\bigstar} \ll \lambda$ 이기만 하면 위의 식을 만족하는 메져 f가 반드시 존재함이 **라돈-니코딤** 정리에 의하여 알려져 있다. 여기에서 $\mu^{\bigstar} \ll \lambda$는 메져 $\mu^{\bigstar}$가 메져 λ에 대하여 **absolutely continuous** 하다는 의미이다(제발 매우작다라고 해석하지 말자). 참고로 원래 **라돈-니코딤**정리를 쓰려면 $\mu^{\bigstar}$, λ가 σ-finite 메져라는 조건이 추가적으로 필요한데 이 경우에는 $\mu^{\bigstar}$는 확률측도이고 λ는 르벡측도이므로 이 조건이 그냥 만족된다. 따라서 증명이 된다.

- 이제 $B := (-\infty, x] \in \mathcal{R}$ 와 $A := X^{-1}(B) \in \mathcal{F}$ 에 대하여 아래와 같은 기묘한 표현들을 파악하여 보자.

$$\begin{aligned}
&\int_A dP = P(A) = P(X \in B) = \int_{X^{-1}(B)} dP = P(X^{-1}(B)) \\
&= \mu^{\bigstar}(B) = \int_B d\mu^{\bigstar} = \int_B \mu^{\bigstar}(dx) = P(X \le x) = \int_{-\infty}^{x} f(\tau)d\tau \\
&= \int_B f d\lambda = \int_B f(x)\lambda(dx) = \int_{-\infty}^{x} dF(\tau) = F(x) - F(-\infty) = F(x)
\end{aligned}$$

참고로 여기에서 B를 일반적인 보렐셋대신 ray를 사용하였는데, 이렇게 쓸 수 있는 이유는 임의의 보렐셋이 $(-\infty, x]$ 들의 **countable union** 으로 쉽게 확장가능하기때문이다. (김김계 해석개론, p295, 명제 10.1.1) 예를들어서 $(1, 3] = (-\infty, 3] - (-\infty, 1]$와 같은 식으로 말이다. 아무튼 이걸 활용하면 임의의

보렐셋에 대한 적분도 표현가능하지만 그 표현법이 지저분하므로 B 를 ray로 단순화 시켰다. 참고로 아래와 같은 표현들도 자주 사용된다.

$$
\begin{aligned}
&\int_A XdP = \int_A X(\omega)P(d\omega) \int_{-\infty}^{x} \tau f(\tau)d\tau \\
&= \int_B xf(x)\lambda(dx) = \int_{-\infty}^{x} \tau dF(\tau) = \int_B xd\mu^\star = \int_B x\mu^\star(dx)
\end{aligned}
$$

참고로 위에서 $\int_{-\infty}^{x} \tau f(\tau)d\tau = \int_{-\infty}^{x} \tau dF(\tau)$ 는 스틸체스 적분의 성질 (김김계 해석개론, p.151, 정리 5.5.3)에 의해서 성립한다.

2 Integration

- 여기에서는 듀렛책 챕터 1.4 $\sim$ 1.7의 내용을 다룬다.

- $f : \Omega \to \mathbb{R}$가 두 가측공간 $(\Omega, \mathcal{F})$와 $(\mathbb{R}, \mathcal{R})$를 이어주는 매저러블-맵핑이라고 하자. 그리고 $\mu : \mathcal{F} \to \mathbb{R}$은 $\mathcal{F}$에서의 메저라고 하자. 여기에서 특별히 μ는 σ-finite measure라고 정의한다. 당분간은 $f = X$ 이고 $\mu = P$ 라고 생각해도 무방하다. (좀 더 발전하면 $f = f(X)$라고 두기도 한다. 어차피 $f(X)$ 즉 function of X는 또 다른 확률변수로 볼 수 있으므로 같은 논의이다.) 이 챕터에서 우리의 관심은 (1) 메저러블-맵핑 f를 메저 μ로 적분한것 즉

$$\int f d\mu$$

를 잘 정의하고 (2) 적분관련 성질들을 자유자재로 계산할 수 있는 적당한 조건을 알아보는 것이다. 여기에서 적분관련 성질들은 예를들어

$$\int af d\mu = a \int f d\mu$$

$$\int f + g d\mu = \int f d\mu + \int g d\mu$$

와 같은 것들을 의미한다.

- $\int f d\mu$를 정의하기 위해 필요한 기본가정을 살펴보자. 여기에서 $\mu := P$ 임을 유의하면서 읽자. (절대로 $\mu = \mu^\star$가 아님)

(1) $f:(\Omega,\mathcal{F})\to(\mathbb{R},\mathcal{R})$ 이 어야 한다. 즉 f 가 메저러블 맵핑 이어야 한다. 왜냐하면 적분 $\int f d\mu$ 를 수행하기 위해서는 임의의 $B\in\mathcal{R}$ 에 대한 f 의 inverse image,

$$A := \{\omega : f \in B\}$$

를 구하고 그 A에 대한 메저 $\mu(A)$를 계산해야 하는데, 이 $\mu(A)$가 정의되기 위해서는 $A\in\mathcal{F}$이어야 하기 때문이다.

(2) μ가 σ-finite measure 이어야 한다. 왜냐하면 이 조건이 있어야 임의의 $A\in\mathcal{F}$에 대하여서도 $\mu(A)<\infty$가 성립하게 되어서 $d\mu$부분을 상당히 안정적으로 정의할 수 있다.

(3) f와 μ가 같은 시그마필드 $\mathcal{F}$를 공유하고 있어야 한다.

(4) 마지막으로 μ와 f는 항상 치역 $\mathbb{R}$을 가져야 한다. μ는 당연히 정의상 치역을 $\mathbb{R}$을 가질수 밖에 없으므로 f의 치역이 항상 $\mathbb{R}$이어야 한다는 조건으로 이해할 수 있다. 즉 f는 치역으로 $\mathbb{R}^n$도 가질 수 없다. 따라서 랜덤벡터와 같이 f의 치역이 $\mathbb{R}^n$인 경우는 바로 적분안된다. (μ의 치역은 $\mathbb{R}$가 되어서 임의의 $A\in\mathcal{F}$에 대하여 $f(A)\mu(A)$를 정의할 수 없다.)

- 왜 적분을 measure theory를 이용하여 정의할까? 고등학교때 정의하던대로 하면 안되나? 다 이유가 있어서 이렇게 정의하는 것이다. measure theory 를 이용하여 적분을 정의하면 고등학교때 배운 논리로 정의하기 애매하 적분들을 매우 깔끔하게 정의할 수 있다는 장점이 있다. 가령 예를들면 모든 유리수에서 1의 값을 가지고 나머지는 0을 가지는 함수 f를 생각하여 보자. 즉 f는 아래와 같다.

$$f(x)=\begin{cases}1 & x\in\mathbb{Q}\\ 0 & x\notin\mathbb{Q}\end{cases}$$

이 함수 f에 대한 적분은 고등학교때 배운 구분구적법 혹은 리만적분과 같은 방식으로는 정의할 수 없다(얼핏 생각하면 리만적분으로는 가능할것 같은데 불가능하다고 한다 f를 리만적분하면 0과 1사이의 임의의 값이 랜덤으로 정의된다고 한다). 하지만 measure theory를 이용하면 이 적분값은 0으로 깔끔하게 정의된다. 즉

$$\int f d\lambda = 0$$

으로 깔끔하게 정의된다. 여기에서 λ는 르벡메져이며 이러한 적분을 **르벡적분**이라고 한다. 기본적으로 구분구적법이나 리만적분은 f가 연속이거나 부분적으로 연속이어야 할 것 같은 느낌인데 measure theory를 이용한 르벡적분은 f가 매저러블 맵핑이기만 하면 되므로 좀 더 적분을 정의할 수 있는 함수의 선택폭이 넓다. 일단 $\mathcal{R}$자체가 거의 대부분의 "subset of $\mathbb{R}$"을 커버하고 있기 때문에 $\mathcal{R}$-메저러블하지 않는 함수 $f : \mathbb{R} \to \mathbb{R}$는 별로없다고 보는 편이 좋다. 즉 르벡적분 불가능한 함수 $f : \mathbb{R} \to \mathbb{R}$는 별로없다고 보는것이 맞다.

• 이제 메저러블-맵핑 f의 적분을 **모순없이** 정의하는 방법을 살펴보자. 적분은 기본적으로 넓이를 구하는 개념이므로 $f : \Omega \to \mathbb{R}$가 **simple function**일때 $\mu : \mathcal{F} \to \mathbb{R}$에 대한 적분은 사각형넓이의 유한합으로 매우 명확히 정의할 수 있다. 이것에 대한 내용이 듀렛책에 설명되어 있지만 당연한 소리라서 생략하겠다.

• 여기까지는 당연한거라서 우리의 관심은 "**f가 simple function이 아닐 경우에도 μ에 대한 적분을 모순없이 잘 정의할 수 있을까?**"가 된다. 결론적으로 **아래의 3가지 경우**에서는 적분값 $\int f d\mu$가 (마치 simple function 인것마냥) 모순없이 잘 정의되며 $\int af d\mu = a \int f d\mu$, $\int f + g d\mu = \int f d\mu + \int g d\mu$ 따위의 성질들도 모순없이 잘 성립한다.

(1) f가 μ 에 대하여 ***거의(=almost) bounded function*** 인 경우

(2) f가 μ 에 대하여 ***거의(=almost) non-negative function***인 경우

(3) f가 μ 에 대하여 ***integrable***인 경우, 즉 $\int |f| d\mu < \infty$ 인 경우

• 먼저 (1) f가 bounded function 인 경우를 살펴보자. f를 두 simple function ϕ, ψ 사이에 끼워넣고, 즉 $\phi \le f \le \psi$ 와 같이 만든 다음 $\int f d\mu$를 아래와 같이 정의하면 f가 simple function 일때 마냥 μ에 대한 적분값을 잘 정의할 수 있다.

$$\sup_{\phi \le f} \int \phi d\mu = \int f d\mu = \inf_{\psi \ge f} \int \psi d\mu.$$

참고로 위의 정의가 말이 되려면 즉 모순이 되지 않으려면

$$\sup_{\phi \le f} \int \phi d\mu = \inf_{\psi \ge f} \int \psi d\mu$$

가 성립해야 하는데 이것이 성립하는 것은 듀렛책에 증명이 되어있다.

• 위의 결과를 이용하면 (2) f가 non-negative function인 경우에도 μ에 대한 적분을 쉽게 정의할 수 있다. 먼저 h를 $0 \le h \le f$를 만족하는 bounded function 이라고 하자. $\int f d\mu$는 아래와 같이 $\int h d\mu$의 sup으로 정의할 수 있다.

$$\int f d\mu = \sup_{0 \le h \le f} \int h d\mu$$

• 이제 (3)의 경우를 살펴보자. 일반적인 f는 아래와 같이 표현가능하다.

$$f = f^+ - f^-$$

따라서 $\int f d\mu$는 아래와 같이 정의할 수 있다.

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu$$

단지 위의 식이 모순없이 정의되려면 $\int f^+ d\mu$와 $\int f^- d\mu$이 모두 무한대이면 안되는데 이런 경우는 $\int |f| d\mu < \infty$라는 조건에 의해서 방지된다.

• 참고로 $\int |f| d\mu < \infty$를 만족하는 f를 우리는 ***integrable하다***고 하는데 이 표현이 좀 사람 헷갈리게 만든다. 즉

$$f \text{ is } \textit{integrable} \text{ wrt to } \mu$$
$$\iff \int f d\mu \text{ has finite value}$$

이다. 그래서 나는 이 표현대신에 ***유한적분값을 가진다***라고 표현하겠다.

note: 즉 교재에서 인테그러블 하다는 것은 말그대로 유한 적분값을 가진다는 말이지 적분을 모순없이 정의할 수 있다는 의미는 아니다. 아래를 참고하라.

- 르벡적분이 불가능한 경우 (= 적분이 well define 되지 않음).
- 르벡적분이 가능한 경우 (= 적분이 well define 됨).
  (i) f가 μ에 대하여 integrable한 경우. 즉 $\int |f| d\mu < \infty$인 경우.
  (ii) f가 non-negative, a.e., wrt μ인 경우. (이 경우 $\int |f| d\mu = \infty$ 가능)
  (iii) f가 bounded, a.e., wrt μ인 경우.

3 Independence

Definition: 독립의 정의는 아래와 같다.

- Two events A and B are indep $\overset{def}{\iff}$ $P(A \cap B) = P(A)P(B)$.
- Two random variables X and Y are indep $\overset{def}{\iff}$ for all $C, D \in \mathcal{R} : P(X \in C, Y \in D) = P(X \in C)P(Y \in D)$.
- Two σ-fields $\mathcal{F}$ and $\mathcal{G}$ are indep $\overset{def}{\iff}$ for all $A \in \mathcal{F}$ and $B \in \mathcal{G}$ the events A and B are indep.

아래예제에서 확인할 수 있듯이 2번째 정의는 3번째 정의의 스페셜케이스이다. [1]

Ex 2.1.1.: (i) X and Y 가 독립이면 $\sigma(X)$ and $\sigma(Y)$ 도 독립임을 보여라. (ii) 역으로 $\mathcal{F}$ and $\mathcal{G}$ 가 독립이고 $X \in \mathcal{F}$ and $Y \in \mathcal{G}$ 이면 X and Y 가 독립임을 보여라.

Ex 2.1.2.: (i) A and B 가 독립이면 A^c and B 도 독립이고 A^c and B^c 도 독립이고 A^c and B^c 도 독립임을 보여라. (ii) 아래를 보여라.

$$A \text{ and } B \text{ are indep} \iff 1_A \text{ and } 1_B \text{ are indep.}$$

- 확률변수의 독립

- 시그마필드의 독립

3.1 Sufficient conditions for Independence

Ex 2.1.18.: 코인을 무한히 던진다고 가정하자. $\Omega = (0,1)$ 이고 $\mathcal{F}$ 를 $(0,1)$ 의 all open interval로 생성한 시그마필드라고 하자. (즉 보렐셋츠.) P 를 르벡메져라고 하자. 확률변수 $Y_n(\omega)$ 를 아래와 같이 정의하자.

$$Y_n(\omega) = \begin{cases} 1 & [2^n\omega] \text{ is odd.} \\ 0 & [2^n\omega] \text{ is even.} \end{cases}$$

그러면 $Y_1, Y_2, \dots$ 는 서로 독립이고 $P(Y_k = 0) = P(Y_k = 1) = 1/2$ 이다.

[1] As the next exercise shows, the second definition is a special case of the third.

4 WLLN

- WLLN 과 SLLN 은 근본적으로 어떠한 확률변수의 합 S_n을 적당한 오더 b_n으로 나누었을때 그 값이 하나의 값으로 p-수렴하는지를 따지는것이다. 즉

$$S_n/b_n \xrightarrow{p} ??$$

혹은

$$S_n/b_n \xrightarrow{as} ??$$

에 대하여 관심이 있다. 가장 이상적으로는 ?? 의 값이 $E(S_n/b_n)$ 이 되는 것인데 이러한 경우를 LLN이 성립한다고 한다. 구체적으로는 p-수렴이 성립하는 경우를 WLLN 이라고 하고 as-수렴이 성립하는 경우를 SLLN 이라고 한다.

- 주의할점은 우리가 항상 $S_n = X_1 + \cdots + X_n$ 이며 $b_n = n$ 인 형태만 관심있는것은 아니라는 것이다. 가령 아래와 같은 것들의 수렴에 관심이 있을 수도 있다.

$$X_1, \frac{X_1 + 2X_2 + X_3}{1+2+1}, \frac{X_1 + 2X_2 + 3X_3 + 2X_4 + X_5}{1+2+3+2+1}, \dots$$

즉 일명 가중평균에 대한 관심이 있을수 있다. 특히 위와 같이 커널과 같은 방식으로 가중치를 줄 경우 수렴성에 대하여 관심을 많이 가진다. 이러한 가중평균을 잘 표현하는 방법중 하나는 바로 삼각수열 $X_{n,1}, \dots, X_{n,n}$ 을 활용하는 경우이다. 여기에서 $X_{n,1}$ 은 **weight** $\times$ X_1로 해석할 수 있다. 따라서 통상적으로 S_n은 아래와 같이 정의하는 것이 맞다.

$$S_n = X_{n,1} + X_{n,2} + \cdots + X_{n,n}$$

- 앞에서 말한것처럼 WLLN 은 아래에 관심이 있다.

$$S_n/b_n \xrightarrow{p} ??$$

여기에서 ?? 의 값은 (case1) 하나의 실수값으로 수렴하지 않을 경우 (case2) 하나의 실수값으로 수렴은 하지만 그 실수값이 $E(S_n/b_n)$은 아닌 경우 (case3) 하나의 실수값으로 수렴하고 그 실수값이 $E(S_n/b_n)$ 인 경우로 나눌 수 있다.

- (case3)이 우리가 원하는 WLLN이다.

- (case1)가 가능하다는것 즉

$$\frac{S_n}{b_n} = \frac{X_{n,1} + \cdots + X_{n,n}}{b_n} \xrightarrow{p} \infty$$

와 같은 경우가 가능하다는것을 이해하는것은 매우 쉽다.

$$X_{n,1} = \text{weight} \times X_1$$

이라고 해석할 수 있는데 weight를 악의적으로 주면 (weight들의 총합이 b_n보다 훨씬 크게) S_n/b_n 을 쉽게 발산시킬 수 있다.

- (case2)가 가능함을 이해하는것은 쉽지않다.

- WLLN에서 전제되는 중요한 조건들을 요약하면 아래와 같다.

(1) 독립 혹은 무상관 조건

(2) 동일한 분포에 대한 조건 (분산만 동일한지, 평균과 분산 모두 같은지, 동일한 분포를 가지는지 등)

(3) 모멘트에 대한 조건 (평균이 존재하는지, 분산이 존재하는지 등)

Lemma 2.2.2: (3) 아래를 만족하는 적당한 r을 선택할 수 있다고 하자.

$$E\left|\frac{X_{n,1} + \cdots + X_{n,n}}{b_n} - \frac{\mu_{n,1} + \cdots + \mu_{n,n}}{b_n}\right|^r \xrightarrow{p} 0$$

그러면 아래와 같은 WLLN이 성립함을 보일 수 있다.

$$\frac{X_{n,1} + \cdots + X_{n,n}}{b_n} \xrightarrow{p} \frac{\mu_{n,1} + \cdots + \mu_{n,n}}{b_n}.$$

Theorem 2.2.3.: (1) $X_1, X_2, \dots$ 이 *uncorrelated random variables*이다. (2) $X_1, X_2, \dots$ 의 평균은 모두 같은 값을 가지지만 분산이 같은 값을 가질필요는 없다. (3) $X_1, X_2, \dots$ 의 분산을 유한한 값 C로 바운드시킬 수 있다. (1)-(3)의 가정하에 아래와 같은 WLLN이 성립한다.

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{p} \mu.$$

note: 참고로 이 조건하에서는 L_2-convergence도 가능하다.

Theorem 2.2.4.: (3) $X_{n,1},\dots,X_{n,n}$ 의 총합의 분산 즉 $V(X_{n,1}+\cdots+X_{n,n})$ 보다 빠르게 발산하는 어떠한 b_n^2 을 잡을 수만 있다면, 즉

$$\frac{V(X_{n,1}+\cdots+X_{n,n})}{b_n^2}\to 0$$

를 만족하는 b_n 을 잡을 수 있다면 아래와 같은 WLLN이 성립한다.

$$\frac{X_{n,1}+\cdots+X_{n,n}}{b_n}\overset{p}{\to}\frac{\mu_{n,1}+\cdots+\mu_{n,n}}{b_n}.$$

note: 조건 $\frac{V(X_{n,1}+\cdots+X_{n,n})}{b_n^2}\to 0$는 결국

$$V(X_{n,1}+\cdots+X_{n,n})=o(b_n^2)$$

이라는 의미이다.

Theorem 2.2.6.: (1) $X_{n,1},\dots,X_{n,k}$ 가 서로 독립이다. (3) $X_{n,1},\dots,X_{n,k}$ 이 아래의 모멘트 조건들을 동시에 만족한다.

$$(\star)\quad \sum_{k=1}^{n}\int \mathbf{1}_{|X_{n,k}|>b_n}dP_k=o(1)$$

$$(\star\star)\quad \sum_{k=1}^{n}\int \frac{1}{b_n^2}X_{n,k}^2\mathbf{1}_{|X_{n,k}|\le b_n}dP_k=o(1)$$

이때 b_n 은 양수이고 무한대로 발산하는 수열이다. (1)-(3)의 가정하에 아래가 성립한다.

$$\frac{X_{n,1}+\cdots+X_{n,n}}{b_n}\overset{p}{\to}\frac{\tilde{\mu}_{n,1}^{b_n}+\cdots+\tilde{\mu}_{n,n}^{b_n}}{b_n}$$

여기에서 $\tilde{\mu}_{n,k}^{b_n}:=E\big(X_{n,k}\mathbf{1}_{|X_{n,k}|\le b_n}\big)=\int X_{n,k}\mathbf{1}_{|X_{n,k}|\le b_n}dP_k$ 이다.

note: 위의 조건 $(\star)$, $(\star\star)$는 Chung의 교재스타일로 정리한것이다. Durret의 노테이션은 아래와 같다.

$$(\star)\quad \sum_{k=1}^{n}P\big(|X_{n,k}|>b_n\big)\to 0$$

$$(\star\star)\quad b_n^{-2}\sum_{k=1}^{n}E\big(X_{n,k}^2\mathbf{1}_{|X_{n,k}|\le b_n}\big)\to 0$$

위에서 엄밀하게 말하면 동일한 P로 쓸수 없고 P_k로 써야한다. E 역시 마찬가지이다. 하지만 간결함을 위해서 (그리고 엄밀함을 희생했다) 듀렛책은 이렇게 표현한것 같다. 대부분 듀렛의 노테이션을 따랐지만 **Thm 2.2.6** 에서만은 Chung의 기호를 빌려왔는데 그 이유는 듀렛책의 표현이 너무 간결하여 의미가 헷갈릴수도 있겠다 싶었기 때문이다.

Theorem 2.2.7.: (1)-(2) $X_1,\dots,X_n$이 ***i.i.d.***라고 하자. (3) $X_1\dots X_n$이 아래의 모멘트 조건을 만족한다고 하자.

$$xP(|X_1|>x)\to 0 \text{ as } x\to\infty \qquad \cdots(\tilde{\star})$$

아래의 WLLN이 성립한다.

$$\frac{X_1+\cdots+X_n}{n}\xrightarrow{p}\tilde{\mu}^n \quad \text{where} \quad \tilde{\mu}^n := E(X_1\mathbf{1}_{|X_1|\le n}).$$

note: **Thm 2.2.7.**은 $\frac{X_1+\cdots+X_n}{n}$ 의 p-수렴성은 보장하지만 그것이 $\mu=EX_1$로 수렴함을 보장하지 않는다. 즉

$$\frac{X_1+\cdots+X_n}{n}\xrightarrow{p}\mu$$

임을 보장하는것은 아니다. 이게 성립하려면 $\tilde{\mu}^n\to\mu$ 가 성립해야 하고 이는 DCT에 의해서 성립해야하는데 DCT가 성립하려면 $E|X_i|<\infty$ 의 조건이 추가적으로 있어야 한다. (어디로인가 수렴한다고 했지 제 정신 박힌 값으로 수렴한다고는 안했다!!) 즉 **Thm 2.2.7**의 조건 아래에서

$$\frac{X_1\mathbf{1}_{|X_1|\le n}+\cdots+X_n\mathbf{1}_{|X_n|\le n}}{n}, \quad \frac{X_1+\cdots+X_n}{n}$$

모두 같은 p-극한을 가진다.

Theorem 2.2.9.: (1)-(2) $X_1,\dots,X_n$이 ***i.i.d.***라고 하자. (3) 유한평균값이 존재한다고 하자. 즉

$$E|X_i|<\infty$$

이라고 하자. 그러면 아래의 WLLN이 성립한다.

$$\frac{X_1+\cdots+X_n}{n}\xrightarrow{p}\mu \quad \text{where} \quad \mu := EX_1.$$

4.1 증명들

Lemma 2.2.2의 증명

- 체비쉐프 부등식을 이용하면 된다.

Theorem 2.2.3의 증명

- L_2 수렴함을 보이면 된다.

$$E\left(\frac{X_1+\cdots+X_n}{n}-\mu\right)\le\frac{Cn}{n^2}=o(1)$$

이므로 증명끝.

Theorem 2.2.4의 증명

- 역시 L_2 수렴함을 보이면 된다.

$$E\left(\frac{X_{n,1}+\cdots+X_{n,n}}{b_n}-\frac{\mu_{n,1}+\cdots+\mu_{n,n}}{b_n}\right)=b_n^{-2}V(X_{n,1}+\cdots+X_{n,n})=o(1)$$

이므로 증명 끝.

Theorem 2.2.6의 증명

- 생략.

Theorem 2.2.7의 증명

- 증명을 위해서 아래의 레마를 외우자. 이걸 증명하는것보다 아래의 레마를 외우는게 더 중요하다.

Lemma 2.2.8.: $Y\ge 0$ 이고 $r>0$이면 아래가 성립한다.

$$E(Y^r)=\int_0^\infty ry^{r-1}P(Y>y)dy.$$

note: 이 정리에서 Y를 양의 확률변수로 한정한 이유는 증명과정에서 푸비니정리를 쓰기 때문이다.

- 조건$(\tilde{\star})$에서 $i.i.d.$를 쓰면 바로 $(\star)$를 얻을수 있다. 우선 조건$(\tilde{\star})$를 살펴보자.

$$(\tilde{\star}):\ \ xP(|X_1|>x)=o(1)\quad \text{as } x\to\infty$$

x 대신에 n을 대입하자.

$$nP(|X_1|>n)=o(1)$$

이제 $i.i.d.$ 조건을 쓰면 아래를 얻을 수 있다.

$$\sum_{k=1}^{n} P\left(|X_k| > n\right) = nP\left(|X_1| > n\right) = o(1)$$

$X_{n,k} = X_k$ 라고 놓고 $b_n = n$ 이라고 두면 이것은 **Thm 2.2.6**의 셋팅에서 $(\star)$를 보인셈이 된다.

- 이제 $(\star\star)$를 보이자. 즉 아래를 보이면 된다.

$$\begin{aligned}
&b_n^{-2}\sum_{k=1}^{n} E\left(X_{n,k}^2 \mathbf{1}_{|X_{n,k}|\le b_n}\right)\\
&= n^{-2}\sum_{k=1}^{n} E\left(X_k^2 \mathbf{1}_{|X_k|\le n}\right) = n^{-2} n E\left(X_1^2 \mathbf{1}_{|X_1|\le n}\right)\\
&= o(1).
\end{aligned}$$

- $Y := X_1 \mathbf{1}_{|X_1|\le n}$ 로 보고 **Lemma 2.2.8**을 쓰자. [2]

$$E\left(X_1^2 \mathbf{1}_{|X_1|\le n}\right) = EY^2 = \int_0^\infty 2yP(Y > y)dy$$

그런데 일단 $Y \le n$ 이므로 $y \ge n$ 이라면 $P(Y > y) = 0$ 이다.[3] 따라서 적분의 범위를 $[0, \infty]$ 에서 $[0, n]$ 으로 좁힐 수 있다. 즉

$$\int_0^{\infty} 2yP(Y > y)dy = \int_0^{n} 2yP(Y > y)dy$$

이다. 아래가 성립함을 관찰하라.

$$\int_0^n 2yP(Y > y)dy = \int_0^n 2yP(X_1 \mathbf{1}_{|X_1|\le n} > y)dy \le \int_0^n 2yP(X_1 > y)dy$$

따라서 아래를 보이면 증명이 끝난다.

$$\frac{1}{n}\int_0^n 2yP(X_1 > y)dy = o(1) \quad \text{or} \quad \int_0^n 2yP(X_1 > y)dy = o(n)$$

이 식은 직관적으로 성립함을 알 수 있다. 왜냐하면

$$2yP(X_1 > y) = o(1) \quad \text{as} \quad y \to \infty$$

[2] $Y := X_1 \mathbf{1}_{|X_1|\le n}$ 이니까 당연히 $Y^2 = X_1^2 \mathbf{1}_{|X_1|\le n}$ 인건 말안해도 괜찮겠지? $\left(\mathbf{1}_{|X_1|\le n}\right)^2 = \mathbf{1}_{|X_1|\le n}$ 이니까?
[3] 잘 모르겠으면 수직선 그려놓고 Y, n, y 순서로 그려봐

이고
$$\frac{1}{n}\int_0^n 2yP(X_1>y)dy$$
는 $2yP(X_1>y)$의 평균같은 개념이기 때문이다.

- 좀 더 엄밀하게 증명하기 위해서[4] 아래를 주목하라.

$$0\le g(y)\le 2y, \quad \text{where} \quad g(y)=2yP(X_1>y)$$

아래가 성립한다.
$$g(y)=o(1)$$
$o(1)$이면 $O(1)$이니까 아래가 성립한다.
$$\sup_y g(y)=M<\infty$$
아래를 정의하자.
$$\epsilon_K=\sup_{y>K} g(y)$$
그리고 $[0,K]$까지의 적분과 $[K,n]$까지의 적분을 따로 생각하자.[5]
$$\int_0^n 2yP(|X_1|>y)dy\le KM+(n-K)\epsilon_K$$
양변을 n으로 나누고 $n\to\infty$를 취하자. 그러면[6]
$$\limsup_{n\to\infty}\frac{1}{n}\int_0^n 2yP(|X_1|>y)dy\le\epsilon_K$$
가 성립한다.[7] K가 arbitary이고 $\epsilon_K\to 0$ as $K\to\infty$이므로 원하는 결론을 얻을 수 있다. [8]

note: 증명의 뒷부분의 논리를 좀 더 내 방식으로 해석해보자. 아래를 정의하자.
$$\epsilon_{K_n}=\sup_{y>K_n} g(y)$$

---
[4] to spell out the details

[5] consider A and B seperately

[6] Dividing by n and letting $n\to\infty$, we have

[7] 굳이 $\limsup$을 쓸 필요는 없어보인다. $\lim$만 써도 충분한듯. 하지만 $\limsup$으로 쓸 수 있어서 그냥 쓴 것 같다.

[8] K가 $K<n$인 조건하에서는 아무렇게나 선택할 수 있다는 의미는 (1) n을 무한대로 보낼때 K도 그에 맞춰서 무한대로 가면서 (2) $K/n\to 0$ as $n\to\infty$ 를 만족하도록 적당히 정할 수 있다는 의미이다. 이는 매우 편리한 성질이다. 풀이과정중 $\frac{KM}{n}$을 0으로 수렴시키게 할 수 있는것은 (2)의 성질을 이용한 것이다. 또한 $\epsilon_K=o(n)$임을 보일때 $\epsilon_K=o(K)$임을 대신 보여도 충분한 것은 (1)의 성질 때문이다.

여기에서 수열 K_n은 아래의 성질을 만족한다.

$$\begin{cases} K_n < n \quad \text{for all} \quad n \\ K_n \to \infty \quad as \quad n \to \infty \\ \frac{K_n}{n} = o(1) \quad as \quad n \to \infty \end{cases}$$

즉 $\{K_n\}$이 n이 무한대로 갈때 그에 맞추어서 점점 무한대로 가는 (하지만 항상 $K_n < n$인) 어떠한 수열이다. [9] 그리고 $[0, K_n]$까지의 적분과 $[K_n, n]$까지의 적분을 따로 생각하자.

$$\int_0^n 2yP(|X_1| > y)dy \le K_n M + (n - K_n)\epsilon_{K_n}$$

양변을 n으로 나누면

$$\frac{1}{n}\int_0^n 2yP(|X_1| > y)dy \le \frac{K_n M}{n} + \epsilon_{K_n} - \frac{1}{n} = \epsilon_{K_n} + o(1) \quad \text{as} \quad n \to \infty$$

와 같이 된다. 그런데

$$\epsilon_{K_n} = o(1) \quad \text{as} \quad n \to \infty$$

이므로 증명이 끝난다.

note: 이게 어심토틱에 나오는 증명의 묘미이다. y에 대한 극한조건을 교묘하게 n 혹은 K 대한 극한조건으로 바꾸어서 증명한다.

Theorem 2.2.9의 증명

- 확률변수가 유한평균값을 가지므로 아래가 성립한다.

$$(\tilde{\star}): \quad xP(|X_1| > x) \to 0 \quad \text{as } x \to \infty$$

참고로 위의 식이 성립하는 이유는

$$xP(|X_1| > x) \le E(|X_1|\mathbf{1}_{|X_1|>x}) \to 0$$

이기 때문이다. 왼쪽의 부등호는 적분꼴로 쓰면 바로 이해가능하고 오른쪽의 극한은 DCT에 의해서 성립한다. 아무튼 조건 $(\star)$ 가 성립하므로 **Thm 2.2.7**을 쓰면

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{p} \tilde{\mu}^n := E(X_1\mathbf{1}_{|X_1| \le n}) \text{ as } n \to \infty$$

[9] 조건 $(\tilde{\star})$에 의해서 $\epsilon_{K_n} = o(1)$ as $n \to \infty$ 가 성립함

가 성립한다. 그런데 아래의 식이 성립한다. (DCT에 의해서 성립함)

$$\tilde{\mu}^n := E(X_1 1_{|X_1|\le n}) \to \mu := EX_1 \text{ as } n \to \infty$$

이제 $S_n/n \xrightarrow{p} \tilde{\mu}^n$ 와 $\tilde{\mu}^n \to \mu$ 임을 이용하면 아래가 성립함을 알 수 있다. ($o_p(1) + o(1) = o_p(1)$ 이므로 성립.)

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{p} \mu.$$

- 증명과정에서 기억할만한 것은 확률변수가 유한평균값을 가진다는 것이 아래를 임플라이 한다는 것이다.

$$(\tilde{\star}): \quad xP(|X_1| > x) \to 0 \quad \text{as } x \to \infty$$

반대로 위 조건(앞으로 편의상 조건 $(\tilde{\star})$ 라고 부르자)이 유한평균값을 가짐을 임플라이 하지는 못한다. 따라서 조건 $(\tilde{\star})$은 유한평균값을 가진다는 조건보다 약한 조건이다. 하지만 조건 $(\tilde{\star})$ 은 확률변수가 유한평균값을 가진다는 조건보다 정말 말 그대로 ϵ 만큼만 약한 조건이다. 구체적으로는 조건 $(\tilde{\star})$ 이 만족하면 아래를 임플라이 할 수 있다.

$$(\tilde{\star}) \quad \Longrightarrow \quad \forall \epsilon > 0: \ E|X_1|^{1-\epsilon} < \infty$$

이것의 증명은 Lemma 2.2.8을 이용해서 할 수 있다. **따라서 조건 $(\tilde{\star})$ 는 유한평균값을 가진다는 조건에 비하여 많이 부족한 조건은 아니다. 사실상 거의 대등하다고 보아도 무방하다.**

note: 조건 $(\tilde{\star})$와 ***uniformly tight***를 헷갈렸었다. (진짜 아는게 없어서 오만걸 다 헷갈리네) 내가 헷갈린 타이트조건은 아래와 같이 쓸 수 있다.

$$\begin{aligned}
&\{X_n\} \text{ is } \textit{uniformly tight} \\
&\overset{def}{\Longleftrightarrow} \forall \epsilon > 0 \quad \exists M \quad s.t. \quad \sup_n P(|X_n| > M) < \epsilon \\
&\overset{def}{\Longleftrightarrow} \sup_n P(|X_n| > M) \to 0 \quad as\ M \to \infty \\
&\overset{def}{\Longleftrightarrow} \sup_n \mu_n^\star\big([-M, M]^c\big) \to 0 \quad as\ M \to \infty \\
&\overset{def}{\Longleftrightarrow} \epsilon > 0 \quad \exists M \quad s.t. \quad \limsup_{n\to\infty} \big(1 - F_n(M) + F_n(-M)\big) < \epsilon \\
&\overset{def}{\Longleftrightarrow} X_n = O_p(1) \quad as\ n \to \infty
\end{aligned}$$

note: 보통 유니폼리-타잇니스도 그냥 줄여서 타잇니스라고 말하기도 한다.

note: 모든 확률변수는 *tight*하다.[10] 하지만 모든 확률변수열이 *uniformly tight* 하지는 않다. [11]

5 Borel-Cantelli Lemmas

5.1 set-theory

- 듀렛교재와 더불어 인터넷을 참고해보자.[12]

(1) $\sup_{k\geq n} A_k = \cup_{k=n}^{\infty} A_k$. [13]

(2) $\inf_{k\geq n} A_k = \cap_{k=n}^{\infty} A_k$.

(3) $\limsup_{n\to\infty} A_n = \lim_{n\to\infty} \sup_{k\geq n} A_k = \lim_{n\to\infty} \cup_{k=n}^{\infty} A_k = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k$. [14]

(4) $\liminf_{n\to\infty} A_n = \lim_{n\to\infty} \inf_{k\geq n} A_k = \lim_{n\to\infty} \cap_{k=n}^{\infty} A_k = \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k$.

- A_n 이 Ω 의 부분집합으로 이루어진 집합열이라고 한다면, 아래와 같이 둘 수 있다. [15]

(1) $\limsup_{n\to\infty} A_n = \Big\{\omega \in \Omega : \sum_{n=1}^{\infty} I(\omega \in A_n) = \infty\Big\} = \Big\{\omega : \omega \in A_n, i.o.\Big\}$

(2) $\liminf_{n\to\infty} A_n = \Big\{\omega \in \Omega : \sum_{n=1}^{\infty} I(\omega \in A_n) < \infty\Big\} = \Big\{\omega : \omega \in A_n, a.b.f.\Big\}$

---

[10] 타이트의 정의는 아래와 같다.

$$
\begin{aligned}
&\text{The random variable } X \text{ is } \textit{tight} \\
&\overset{def}{\Longleftrightarrow} \forall \epsilon > 0 \quad \exists M \quad s.t. \quad P(|X| > M) < \epsilon \\
&\overset{def}{\Longleftrightarrow} P(|X| > M) \to 0 \quad as\ M \to \infty \\
&\overset{def}{\Longleftrightarrow} \mu^\star\big([-M, M]^c\big) \to 0 \quad as\ M \to \infty \\
&\overset{def}{\Longleftrightarrow} \epsilon > 0 \quad \exists M \quad s.t. \quad 1 - F(M) + F(-M) < \epsilon
\end{aligned}
$$

[11] 확률변수열이 *uniformly tight* 하다는 말은 확률변수열이 유계라는 말과 같다. Van der Vaart (2000)의 p.8.

[12] [1] `http://theanalysisofdata.com/probability/A_4.html`, [2] `https://en.wikipedia.org/wiki/Set-theoretic_limit`

[13] 여기에서 $\sup_{k\geq n} A_k$ 는 sequence of n 임을 유의하자. 즉 $\sup_{k\geq n} A_k = B_n$ 과 같이 놓을 수 있다. 하지만 k 에 대한 sequence는 아니기 때문에 $\sup_{k\geq n} A_k = B_k$ 와 같이 놓을 수는 없다.

[14] 마지막등호는 $\cup_{k=n}^{\infty} A_k := B_n$ 이 n 에 대한 monotone sequence이기 때문에 성립한다.

[15] If A_n is subset of sequence of Ω, we let

example: $X_1, X_2, \ldots$ 이 베르누이열이라고 하자. [16] 그리고 $(\Omega_1, \mathcal{F}_1, P_1)$ 을 X_1 에 대응하는 확률공간이라고 하자. 콜모고로프정리에 따라서 랜덤벡터 $(X_1, X_2, \ldots,)$가 정의되는 공간 $(\Omega, \mathcal{F})$를 정의할 수 있다. [17] 이제 $S_n/b_n = \frac{1}{n}\sum_{i=1}^n X_i$ 라고 하자. 이제

$$A_n = \Big\{ \omega \in \Omega : |S_n(\omega)/b_n - 1/2| > \epsilon \Big\}$$

라고 정의하자. [18] 그러면

$$\sup_{\epsilon>0}\Big\{ \limsup_{n\to\infty} A_n \Big\} = \Big\{ 0\text{보다 } 1\text{의 숫자가 } \textcolor{red}{\text{일정한 비율로}} \text{ 많은 모든 이벤트} \Big\}$$

이다.

note: $\sup_{\epsilon>0}\Big\{ \limsup_{n\to\infty} A_n \Big\}$의 원소로 어떤것이 있을지 알아보자. 빨간색글자는 copy and paste 로 반복된다고 하자.

$$\begin{aligned}
(1):&\quad \{1, \mid \textcolor{red}{1,0, \mid 1,0, \mid ,\ldots}\} \\
(2):&\quad \{1,1, \mid \textcolor{red}{1,0, \mid 1,0, \mid ,\ldots}\} \\
(3):&\quad \{1,1,1, \mid \textcolor{red}{1,0, \mid 1,0, \mid ,\ldots}\} \\
(4):&\quad \{\textcolor{red}{1, \overbrace{1,0,1,0,1,0,1,0,\ldots,1,0}^{\times 100}, \mid 1, \overbrace{1,0,1,0,1,0,1,0,\ldots,1,0}^{\times 100}, \mid ,\ldots}\}
\end{aligned}$$

(1)-(3)의 경우는 $\sup_{\epsilon>0}\Big\{ \limsup_{n\to\infty} A_n \Big\}$의 원소이지만 (4)의 경우는 아니다.[19]

6 Poisson Convergence

Theorem: Each n 에 대하여 $X_{n,m}$, $1 \le m \le n$ 이 서로 독립인 확률변수라고 하자. 그리고 $P(X_{n,m}=1) = p_{n,m}$ 이고 $P(X_{n,m}=0) = 1-p_{n,m}$이라고 하자. (1) $\sum_{m=1}^n p_{n,m} \to \lambda$ *as* $n \to \infty$ (2) $\max_{1\le m\le n} p_{n,m} \to 0$ *as* $n \to \infty$ 이라면 아래가 성립한다.

$$S_n \Rightarrow Z \quad \text{where } Z \text{ is Poisson}(\lambda).$$

[16] 즉 $X_1, X_2, \ldots$ 의 가능한 realization은 $0, 1, \ldots$ 이다.

[17] 즉 $\Omega := \Omega_1 \times \Omega_2 \times \ldots$ 이고 $\mathcal{F} := \mathcal{F}_1 \times \mathcal{F}_2 \times \ldots$.

[18] A_n은 ϵ에 따라 변화하는 함수인데 이를 강조하기 위해서 $A_n(\epsilon)$이라고 표시할 수도 있다.

[19] 왜냐하면 (4)의 경우는 $\epsilon = \frac{1}{101}$ 보다 크면 $\limsup_{n\to\infty} A_n$ 의 원소이지만 $\epsilon = \frac{1}{101}$ 보다 작으면 $\limsup_{n\to\infty} A_n$ 의 원소가 아니기 때문이다.

note: 이항분포의 포아송근사. 일단 독립인 랜덤변수들의 합의 분포를 다루는 것임.

6.1 Two Example with Dependence

- 독립이 아닌경우에서도 포아송으로 근사할 수 있다는 예제들을 설명한 챕터인 듯.

7 Limit Theorems in $\mathbb{R}^d$

exercise: $F_1, \ldots, F_d$가 $\mathbb{R}$에서 정의되는 distribution이라고 하자. [20] 그러면 임의의 $\alpha = [-1, 1]$에 대하여

$$F(x_1, \ldots, x_d) = \left\{1 + \alpha \prod_{i=1}^{d}(1 - F_i(x_i))\right\} \prod_{j=1}^{d} F_j(x_j)$$

역시 d.f. 이고 $F(x_1, \ldots, x_d)$의 마지날은 각각 $F_1, \ldots, F_d$와 같다. 이때 $\alpha = 0$은 $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ 의 각 원소가 독립인 경우를 의미한다.

[20] Let $F_1, \ldots, F_d$ be the distributions on $\mathbb{R}$.