4.1. Stopping Times

- 아래를 가정하자.

$$\begin{aligned}&\Omega = \{(\omega_1, \omega_2, \dots,) : \omega_i \in S\}\\&\mathcal{F} = \mathcal{S} \times \mathcal{S} \times \dots\\&P = \mu \times \mu \times \dots \quad \mu \text{ is the distribution of } X_i.\\&X_n(\omega) = \omega_n\end{aligned}$$

- 맵핑 $\pi : \mathbb{N} \to \mathbb{N}$ 가 *finite permutation*이라는 것은 유한개의 i에 대해서면 $\pi(i) \neq i$가 성립한다는 의미이다. 따라서 퍼퓨테이션은 인덱스의 순서를 (유한개) 바꿔주는 함수라 볼 수 있다.

- 만약에 π가 $\mathbb{N}$의 *finite permutation*이고 $\omega \in S^{\mathbb{N}}$이라면 아래와 같이 정의할 수 있다.

$$i\text{-th element of } \pi\omega = \omega\text{의 } \pi(i)\text{-th element}$$

따라서

$$(\pi\omega)_i = \omega_{\pi(i)}$$

와 같이 쓸 수 있다. 이 경우 array of realization:

$$\omega_1, \omega_2, \dots$$

이 π에 의해서 재정렬된다고 해석할 수 있다. (Note that $\omega_1 = X_1(\omega)$)

- 따라서 array of realization 즉 어떠한 확률시행의 결과에 π를 취한다는 의미는 확률변수의 결과 중 유한개의 순서를 임의로 바꾼다는 것을 의미한다.

(def) 어떠한 사건 A가 *permutable*하다는 것은 사건 A가 아래를 식을 만족한다는 의미이다.

$$A = \pi^{-1}(A)$$

여기에서 $\pi^{-1}(A) = \{\omega : \pi\omega \in A\}$이다. 이 정의는 이해하기 그렇게 쉽지 않다. 아래의 설명들을 참고하여 보자.

- 우선 사건 A란 무엇인지 다시 생각해보자. 사건이란 확률시행의 결과를 재해석하여 구성한 어떠한 이벤트를 의미한다. 따라서 A는 예를들어 "주사위를 5번 던졌을때 짝수가 3번이상 나올 사건"과 같이 정의할 수 있다.

- 이제 퍼뮤터블의 의미를 살펴보자. 어떠한 사건 A 퍼뮤터블하다는 것은 확률변수의 결과 중 유한개의 순서를 임의로 바꾸어도 사건 A를 일관적으로 정의할 수 있다는 것을 의미한다.

example: 예를들어 보자. $X_1, X_2, \dots$ 이 1과 -1중 하나가 나오는 베르누이 시행이라고 하자. 만약

$$A : \text{랜덤변수들의 총합 즉 } X_1 + X_2 + X_3 + \dots \text{ 가 음수일 사건}$$

이라 정의한다고 하자. 직관적으로 A는 확률변수들의 순서를 바꾸어서 상관없으므로 이럴 경우 사건 A를 퍼뮤터블 하다고 말한다. 이 예제를 좀더 수학적으로 표현하여 보자.

$$A = \{\omega : X_1(\omega) + X_2(\omega) + \dots < 0\}$$

π를 인덱스 1과 2를 바꾸는 변환이라고 하자. 그러면

$$\pi\omega = \pi(\omega_1, \omega_2, \dots) = (\omega_2, \omega_1, \dots)$$

따라서

$$\omega \in A \iff X_1(\omega) + X_2(\omega) + X_3(\omega) + \dots < 0$$

에 대응하는 것은

$$\pi\omega \in A \iff X_2(\omega) + X_1(\omega) + X_3(\omega) + \dots < 0$$

결국

$$\pi^{-1}(A) = \{\omega : \pi\omega \in A\} = \{\omega : \omega \in A\} = A.$$

example: 편의상 X_1, X_2, X_3을 베르누이시행에서 관측하였다고 하자.

$$A\text{: } X_1 + X_2 + X_3 = 1 \text{ 일 사건}$$

그러면

$$A = \{\omega : \boldsymbol{X}(\omega) = (1,0,0) \text{ or } \boldsymbol{X}(\omega) = (0,1,0) \text{ or } \boldsymbol{X}(\omega) = (0,0,1)\}$$

이다. 단, $\boldsymbol{X}(\omega) = \big(X_1(\omega), X_2(\omega), X_3(\omega)\big)$. 이때 A는 퍼뮤터블하다. 아래와 같은 collection을 고려하여 보자.

$$\mathcal{E} = \{\emptyset, A, A^c, \Omega\}$$

$\mathcal{E}$는 시그마필드이고 모든 원소가 퍼뮤터블하다.

(def) $\mathcal{E}$를 *exchangeable* 시그마필드라고 말한다. 즉 어떠한 시그마필드를 구성하는 모든 사건이 퍼뮤터블하면 익스체인지어블하다고 말한다.

example: 퍼뮤터블이벤트에 대한 예제를 좀더 살펴보자. (교재에 있는 예제이다.) 아래의 사건은 퍼뮤터블하다.

$$\{\omega : S_n(\omega) \in B,\ i.o.\}$$

왜냐하면 적당히 큰 n에 대하여 $S_n(\omega) = S_n(\pi\omega)$이 성립하기 때문이다.

note: 참고로 아래표현을 기억하면 위의 state를 이해하기에 좋다.

$$\begin{aligned}&S_n \in B,\ i.o.\\&\iff S_n \in B^c,\ a.b.f.\\&\iff S_n \in B^c,\ \text{for all } n \geq N_0\end{aligned}$$

note: $S_n(\omega) = S_n(\pi\omega)$은 그냥 모든 n에 대하여 성립하는것 아닌가? 하는 착각을 하지 않기를 바란다. 물론

$$S = X_1 + X_2 + X_3 + \dots$$

은 항상

$$S(\omega) = S(\pi\omega)$$

가 성립하지만 S_n은 그렇지 않다. 예를들어서 π를 1과 100의 인덱스를 서로 바꾸는 규칙이라고 하자.

$$\begin{aligned}&S_1(\omega) = X_1 \text{ and } S_1(\pi\omega) = X_{100}\\&S_2(\omega) = X_1 + X_2 \text{ and } S_2(\pi\omega) = X_{100} + X_2\\&\dots\end{aligned}$$

이므로 $n < 100$에 대하여서는 $S_n(\omega) \neq S_n(\pi\omega)$이다.

note: 여기에서 적당히 큰 n에 대하여

$$S_n(\omega) = S_n(\pi\omega)$$

임을 주장할 수 있다. 즉 적당히 큰 수 N_0가 존재하여 $n \geq N_0$인 모든 n에 대하여

$$S_n(\omega) = S_n(\pi\omega)$$

가 성립함을 보일 수 있다. 그도 그럴것이 π는 유한개의 인덱스를 서로 바꾸는 역할을 하므로 적당히 큰 수 N_0이상으로는 그 유한개의 바꿈이 없다고 보일수 있다. 이때 큰 수 N_0의 존재성은 "유한"개의 바꿈이라는 조건때문에 증명된다. 예를들어 100을 1로 1을 100으로 바꾼다고 하자. 그러면 $M = 100$이 존재하여 $n \geq N_0$에서는 모두

$$S_n(\omega) = S_n(\pi\omega)$$

라고 주장할 수 있다.

example: 아래의 사건도 퍼뮤터블하다.

$$\left\{\omega : \limsup_{n\to\infty} S_n(\omega)/c_n \leq 1\right\}$$

이유는 적당히 큰 n에 대하여 $S_n(\omega) = S_n(\pi\omega)$라고 주장할 수 있기 때문이다.

(thm) 테일-시그마필드에 속하는 모든 이벤트는 퍼뮤터블하다. 이때 *tail-σ-field*는 아래와 같이 정의되는 시그마필드 $\mathcal{T}$ 이다.

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{F}'_n \quad \text{where } \mathcal{F}'_n = \sigma(X_n, X_{n+1}, X_{n+2}, \dots)$$

note: 어떠한 이벤트 A가 $A \in \mathcal{T}$라는 것은 임의의 유한개의 realization의 결과를 완전히 바꾸어도 event가 동일함을 의미한다. 즉 X_1, X_{100}의 순서만 바꾸는 것이 아니고 X_1의 값과 X_{100}을 아예 임의의 다른값으로 바꾸어도 된다는 의미이다.

- 위의 정리의 역은 성립하지 않는다. 즉 퍼뮤터블한 사건이 항상 테일 시그마 필드의 원소는 아니다. 아까 소개한 바 있는 아래의 사건

$$\{\omega : S_n(\omega) \in B,\ i.o.\}$$

은 $\mathcal{E}$에 속하지만(=퍼뮤터블하지만) $\mathcal{T}$에 속하지 않는(=테일시그마필드의 원소는아닌) 사건이다. (왜???)

- 만약에 확률변수열 $X_1, X_2, \dots$ 이 iid 라면 $\mathcal{E}$와 $\mathcal{T}$는 차이가 없다. 이게 바로 **휴이트-새비지**의 정리이다.

(Theorem 4.1.1. Hewitt-Savage 0-1 law.) (1) $X_1, X_2, \dots$ 이 i.i.d.이고 (2) $A \in \mathcal{E}$ 이라면

$$P(A) \in \{0, 1\}$$

이다.

6.1. Definitions

(def) $(S, \mathcal{S})$를 measurable space라고 하자. $X_n : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$이라고 하자. 편하게

$$(S, \mathcal{S}) = (\mathbb{R}, \mathcal{R})$$

이라고 생각해도 무방하다. 어떠한 확률변수열 $\{X_n\}$이 filtration $\mathcal{F}_n := \sigma(X_0, \dots, X_n)$에서 정의되어 있다고 하자. 확률변수열 $\{X_n\}$이 마코프체인이라는 것은 아래와 같이 정의한다.

$$\begin{aligned}&\{X_n\} \text{ is Markovchain w.r.t. } \mathcal{F}_n\\&\overset{def}{\iff} \text{for all } B \in \mathcal{S}: \quad P(X_{n+1} \in B | \mathcal{F}_n) = P(X_{n+1} \in B | X_n)\end{aligned}$$

note: 확률변수열 $X_1, X_2, \dots$ 의 값이 바로 이전의 값에 의해서만 결정되면 마코프체인이라고 한다. 즉 X_2의 값을 알기 위해서는 X_1의 값에 대한 정보만 있으면 되고 X_3의 값을 알기 위해서는 X_2에 대한 정보만 있으면 될때 $X_1, X_2, \dots$ 을 마코프체인이라고 한다.

- 4×4 그리드 세계를 가정하자.

$$\Omega = \{(1,1), \dots, (4,4)\}$$

이고

$$S = \{1, 2, 3 \dots, 16\}$$

이라고 하자. 확률변수 $X_1 : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$은 아래와 같이 정의할 수 있는 맵핑이라고 하자.

$$\begin{aligned}&X_1\big((1,1)\big) = 1\\&X_2\big((1,2)\big) = 2\\&\dots\\&X_1\big((4,4)\big) = 16\end{aligned}$$

따라서

$$X_1 = 1, X_2 = 2$$

가 의미하는 것은 처음에는 $(1,1)$의 위치에 있다가 그다음에는 $(1,2)$의 위치로 이동하였다는 것을 의미한다. 이제 $(1,1)$의 위치에서 $(1,2)$의 위치로 이동하는 transition probability를 p라고 정의하자. 여기에서 p는 확률공간을 구성하는 P이지 pmf를 의미하는 것이 아님을 기억하자. p는 아래와 같이 정의할 수 있다.

$$p : (S, \mathcal{S}) \to \mathbb{R}$$

기호로는 아래와 같이 쓴다.

$$p(x, A)$$

여기에서 $x \in S$, $A \in \mathcal{S}$이다.

example: $x = 1$, $A = \{1, 2, 5\}$이라고 하자.

$$\begin{aligned}&x = 1 \Leftrightarrow X(\omega) = 1 \Leftrightarrow \omega = (1,1)\\&A = \{1,2,5\} \Leftrightarrow \{\omega : X(\omega) \in A\} = \{(1,1), (1,2), (2,1)\}\end{aligned}$$

임을 주목하라. 따라서

$$p(x, A)$$

는 점 $(1,1)$에서 출발했는데 점 $(1,1), (1,2), (2,1)$중 하나에 도착할 확률이므로

$$p(x, A) = 1$$

이라고 볼 수 있다.

(def) transition probability의 정의를 사용하면 아래를 만족하는 확률변수열 $\{X_n\}$을 마코프체인이라 정의할 수 있다.

$$P(X_{n+1} \in B | \mathcal{F}_n) = p_n(X_n, B)$$

여기에서 p_n은 n번째에 어떠한 위치 X_n에서 B의 부분집합중 하나의 위치로 이동할 확률을 의미한다.

(결론1,2,3의 가정) 만약에 (1) $(S, \mathcal{S})$이 *nice space* 이고 (2) $\{p_n\}$이 잘 정의되며 (3) $(S, \mathcal{S})$에서의 initional distribution μ가 잘 정의된다고 하자.

(결론1) 일단 유한개의 확률변수열 $\{X_n\}$에 대하여 consistence set of finite dimensional distribution을 아래와 같이 잘 정의할 수 있다.

$$\begin{aligned}&\text{Prob}(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n)\\&= P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n)\\&= \int_{B_0} \mu(dx_0) \int_{B_1} p_0\big(x_0, \mu(dx_1)\big) \cdots \int_{B_n} p_{n-1}\big(x_{n-1}, \mu(dx_n)\big)\end{aligned}$$

note: 이때 $\text{Prob}(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n)$와 같은 표현은 이해하기 쉽지만 수학적으로 엄밀하지 않은 표현이다. 따라서 엄밀하게 하려면 아래의 공간에서 정의되는 확률측도 P를 사용하여 표현해야한다.

$$(S_0 \times S_1 \cdots \times S_n, \mathcal{S}_0 \times \mathcal{S}_1 \cdots \times \mathcal{S}_n)$$

이 공간은 간단하게 아래와 같이 표현하기도 한다.

$$\left(S^{\{0,1,\dots,n\}}, \mathcal{S}^{\{0,1,\dots,n\}}\right)$$

note: 즉 결론1은 $\big(S^{\{0,1,\dots,n\}}, \mathcal{S}^{\{0,1,\dots,n\}}\big)$에서의 확률측도 P는 (혹은 임의의 유한 확률변수열 $\{X_n\}$에 대한 확률측도 P는) 초기분포 μ와 p_n만 잘 정의하면 모순없이 정의가능하다는 것을 의미한다.

- 하지만 무한일 경우에도 잘 정의될까?

(결론2) (1)-(3)의 가정하에 Kolmogorov's theorem은 확률변수열 $\{X_n\}$이 무한수열을 가지더라도 아래와 같은 확률이 잘 정의됨을 보여준다.

$$\text{Prob}(X_1 \in B_1, X_2 \in B_2, \dots,)$$

즉 이는 μ와 $\{p_n\}$만 잘 정의되면 위의 같은 확률들을 모순없이 정의할 수 있음을 의미한다. 교재에서는 유한인 경우와 구분하기 위해서 위의 확률을 표현하는 확률측도를 특별히 P_μ라고 하였다. 즉 아래와 같이 써야 올바르다.

$$P_\mu(X_1 \in B_1, X_2 \in B_2, \dots)$$

이때 P_μ는 $\big(S^{\{0,1,\dots\}}, \mathcal{S}^{\{0,1,\dots,\}}\big)$에서의 확률측도이다.

note: 아래의 기호는 외우는 것이 좋겠다.

$$\begin{aligned}&\text{Prob}(X_0 \in B_0) = \int_{B_0} \mu(dx_0)\\&\text{Prob}(X_0 \in B_0, X_1 \in B_1) = \int_{B_0} \mu(d(x_0)) \int_{B_1} p\big(x_0, \mu(dx_1)\big)\end{aligned}$$

- 지금까지는 콜모고로프의 정리덕에 P_μ가 잘 정의된다는 사실까지 살펴보았다. 즉 (1) 초기분포 μ와 (2) transition 확률 $\{p_n\}$이 잘 정의되면 무한하게 눈을 쌓아도 P_μ가 잘 정의된다.

(결론3, Thm 6.1.1) (1)-(3)의 조건하에 $\{X_n\}$이 마코프체인이 된다.

(pf) 아래의 기호를 정의하면서 증명을 시작하자.

(notation) $\mu = \delta_x$를 x에서의 포인트매스라고 하자. 그리고 기호 $P_x = P_{\delta_x}$라고 정의하자. P_x가 정의되면 아래와 같이 P_μ를 정의할 수 있다.

$$P_\mu(A) = \int \mu(dx) P_x(A), \quad A \in \mathcal{S}^{\{0,1,\dots,\}}$$

(Thm 6.1.2, 결론1의 변형) (1)-(3)의 조건중 체크하기 까다로운 것은 (1)이다. 오히려 (1)의 조건대신에 $\{X_n\}$이 마코프체인임을 가정하면 결론2와 동일한 결과를 얻을 수 있다. 즉 (1) $\{X_n\}$이 마코프체인이고 (2) transition prob $\{p_n\}$이 주어졌고 (3) initional distribution μ 가 주어졌다면 *finite dimensional distribution*이 아래와 같이 주어진다.

$$\begin{aligned}&P(X_j \in B_j, 0 \leq j \leq n)\\&= P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n)\\&= \int_{B_0} \mu(dx_0) \int_{B_1} p_0\big(x_0, \mu(dx_1)\big) \cdots \int_{B_n} p_{n-1}\big(x_{n-1}, \mu(dx_n)\big)\end{aligned}$$

6.2. Examples

6.3. Extensions of the Markov Property

6.4. Recurrence and Transience

6.5. Stationary Measures

- 아래식을 만족하는 measure μ를 stationary measure라고 한다.

$$\sum_x \mu(x) p(x, y) = \mu(y)$$

note: $p(x, y)$: 노드 x에서 다음 노드 y로 이동할 확률

note: $\mu(x)$: 노드 x에 있을 확률

note: 따라서 stationary measure는 특정노드에 있을 확률을 측정하는 메저라 생각할 수 있다.

- stationary measure(=stationary distribution)가 (1) 존재하고 (2) 유일하다는 것이 조사되었다고 하자. 이제 다음 관심사는 아래식을 만족하는 staionary distribution π이다.

$$\pi p = \pi$$

(정리 6.5.6.)

(정리 6.5.6.) p가 irreducible 하다는 것과 아래는 동치이다.

(1) .

(2) stationary distribution이 존재한다.

(3) .

Asymptotic Behavior

(레마 6.6.3.) $d_x = 1$이라면 m_0보다 큰 모든 m에 대하여

$$p^m(x, x) > 0$$

를 만족시킬 수 있다.

(정리 6.6.4.) p가 (1) irreducible 하고 (2) aperiodic 하며 (3) stationary distribution π를 가진다고 하자. 그러면 아래가 성립한다.

$$p^n(x, y) \to \pi(y) \quad as\ n \to \infty.$$

note: p가 irreducible 인것만 보이면 stationary distribution π를 가진다는 것은 정리 6.5.6에 의해서 성립한다. 따라서 (1)-(2)만 조건으로 사용해도 위의 정리는 성립한다.

note: p가 에이피리오딕하다는 의미는 모든 state가 $d_x = 1$을 가진다는 것을 의미한다 .

(pf)

- $S^2 = S \times S$ 라고 하자.

- 전이확률 $\bar{p}$를 $S \times S$에서 아래와 같이 정의하자.

$$\bar{p}\big((x_1, y_1), (x_2, y_2)\big) = p(x_1, x_2)p(y_1, y_2)$$

note: 이는 각각의 coordinate가 독립적으로 움직인다는 것을 의미한다.

(step 1)

- 먼저 $\bar{p}$가 이리듀시블임을 보이자. (이는 너무 당연해서 바보같은 증명으로 보이지만 정리의 에이피리오딕조건을 사용하는 유일한 과정이다.) 우선 p가 이리듀시블하다는 조건으로부터 아래를 만족하는 적당한 K, L이 존재함을 알 수 있다.

$$p^K(x_1, x_2) > 0 \quad \text{and} \quad p^L(y_1, y_2) > 0.$$

그런데 레마 6.6.3에 의해서 M을 적당히 크게 설정한다면 아래를 만족시킬수 있음을 알 수 있다.

$$p^{L+M}(x_1, x_2) > 0 \quad \text{and} \quad p^{K+M}(y_1, y_2) > 0.$$

따라서 아래가 성립한다.

$$\bar{p}^{K+L+M}\big((x_1, y_1), (x_2, y_2)\big) > 0$$

(step 2)

- 두 코디네이츠가 독립이므로 $\bar{p}$의 stationary distribution을 아래와 같이 정의할 수 있다.

$$\bar{\pi}(a, b) = \pi(a)\pi(b).$$

- 정리 6.5.4에 의해서 $\bar{p}$의 stationary distribution이 존재한다는 것은 $\bar{p}$의 모든상태가 recurrent하다는 것을 의미한다.

- (X_n, Y_n)을 $S \times S$에서의 체인이라고 하자.

- T를 이 체인이 처음으로 대각 $\{(y, y) \in S\}$을 치는 시간이라고 하자.

- $T(x, x)$를 (x, x)를 hit하는 시간이라고하자.

- $\bar{p}$가 (1) irreducible 하고 (2) recurrent 하므로 $T(x, x) < \infty\ a.s.$ 이고 따라서 $T < \infty\ a.s.$ 이다.

(step 3)

- 우선 두개의 코디네이트 (X_n, Y_n)가 $\{T \le n\}$에서 같은 분포를 가진다는 것을 관찰하자.

- (X_n, Y_n)이 첫 교차점을 가지는 시간과 장소를 고려하여보자. 마코프 성질을 이용하면

$$P(X_n = y, T \le n) = \sum_{m=1}^{n} \sum_{x} P(T = m, X_m = x, X_n = y)$$

6.6. Periodicity, Tail σ-field

6.7. General State