Probability
(확률론)

Fifth Edition · Version 1.0

Author : Rick Durrett
Translated by Dyne

Nov 02, 2023

Contents (목차)

1. Measure Theory (측도론) **3**

2. Laws of Large Numbers (큰 수의 법칙) **4**

2.1 Independence (독립성) . . . 4

2.1.1 Sufficient Conditions for Independence (독립성의 충분조건) . . . 5

2.1.2 Independence, Distribution, and Expectation (독립성, 확률분포, 기댓값) . . . 7

2.1.3 Sums of Independent Random Variables (독립 확률변수들의 합) . . . 8

2.1.4 Constructing Independent Random Variables (독립 확률변수의 구축) . . . 10

2.2 Weak Laws of Large Numbers (대수의 약법칙) . . . 12

2.2.1 L^2 Weak Laws (L^2 대수의 약법칙) . . . 12

3. Central Limit Theorems (중심극한정리) **14**

4. Martingales (마팅게일) **15**

4.1 Conditional Expectation (조건부기댓값) . . . 15

4.1.1 Examples (예시) . . . 16

4.1.2 Properties (성질) . . . 18

4.1.3 Regular Conditional Probabilities (정칙조건부확률)* . . . 21

4.2 Martingales, Almost Sure Convergence (마팅게일, 거의 확실한 수렴) . . . 21

4.3 Examples (예시) . . . 24

4.3.1 Bounded Increments (유계 증가) . . . 25

4.3.2 Polya's Urn Scheme (Polya 항아리 구조) . . . 26

4.3.3 Radon-Nikodym Derivatives (Radon-Nikodym 도함수) . . . 27

4.3.4 Branching Processes (분기 과정) . . . 29

4.4 Doob's Inequality, Convergence in $L^p, p > 1$ (Doob 부등식, $p > 1$인 L^p에서의 수렴) . . . 31

4.5 Square Integrable Martingales (제곱 적분가능 마팅게일)* . . . 33

4.6 Uniform Integrability, Convergence in L^1 (균등적분가능성, L^1 수렴성) . . . 33

4.7 Backwards Martingales (후진 마팅게일) . . . 37

4.8 Optimal Stopping Problem (최적 정지 문제) . . . 39

4.8.1 Applications to Random Walks (난보에의 응용) . . . 40

1 | Measure Theory (측도론)

2 Laws of Large Numbers (큰 수의 법칙)

2.1 Independence (독립성)

독립성의 정의에서 측도론이 끝나고 확률론이 시작된다. 우리는 익숙한 정의에서 시작하여 현재 설정에 적합한 정의로 진행할 것이다.

두 사건 A, B가 **독립(independent)**임은 $P(A \cap B) = P(A)P(B)$인 것이다.
두 확률변수 X, Y가 **독립(independent)**임은 모든 $C, D \in \mathcal{R}$에 대하여 다음이 성립하는 것이다.

$$P(X \in C, Y \in D) = P(X \in C)P(Y \in D)$$

i.e. 사건 $A = \{X \in C\}$와 $B = \{Y \in D\}$가 독립이다.
두 σ-체 $\mathcal{F}, \mathcal{G}$가 **독립(independent)**임은 모든 $A \in \mathcal{F}$와 $B \in \mathcal{G}$에 대하여 사건 A, B가 독립인 것이다.

다음 결과는 둘째 정의가 셋째 정의의 특수한 경우임을 보여준다.

Theorem 2.1.1.
(i) 만약 X와 Y가 독립이면 $\sigma(X)$와 $\sigma(Y)$가 독립이다.
(ii) 역으로 만약 $\mathcal{F}, \mathcal{G}$가 독립이며 $X \in \mathcal{F}, Y \in \mathcal{G}$이면 X, Y가 독립이다.

Proof. (i) 만약 $A \in \sigma(X)$이면 $\sigma(X)$의 정의에서 어떠한 $C \in \mathcal{R}$에 대하여 $A = \{X \in C\}$임이 따라온다. 마찬가지로 만약 $B \in \sigma(Y)$이면 어떠한 $D \in \mathcal{R}$에 대하여 $B = \{Y \in D\}$이다. 그러므로 이러한 사실들과 X, Y의 독립성을 이용하면 다음이 성립한다.

$$P(A \cap B) = P(X \in C, Y \in D) = P(X \in C)P(Y \in D) = P(A)P(B)$$

(ii) 역으로 만약 $X \in \mathcal{F}, Y \in \mathcal{G}, C, D \in \mathcal{R}$이면 가측성의 정의로부터 $\{X \in C\} \in \mathcal{F}, \{Y \in D\} \in \mathcal{G}$임이 따라온다. $\mathcal{F}, \mathcal{G}$가 독립이므로 $P(X \in C, Y \in D) = P(X \in C)P(Y \in D)$이다. □

첫째 정의는 다시 둘째 정의의 특수한 경우이다.

Theorem 2.1.2.
(i) 만약 A와 B가 독립이면 A^c와 B, A와 B^c, A^c와 B^c도 독립이다.
(ii) 사건 A와 B가 독립일 필요충분조건은 지시확률변수 1_A와 1_B가 독립인 것이다.

Proof. (i) $P(A \cap B) = P(A)P(B)$를 $P(B) = P(B)$에서 빼면 $P(A^c \cap B) = P(A^c)P(B)$임을 얻는다. 둘째와 셋째 결과는 첫째 결과를 독립 사건들의 쌍 (B, A)와 (A, B^c)에 적용하면 따라온다.
(ii) 만약 $C, D \in \mathcal{R}$이면 $\{1_A \in C\} \in \{\emptyset, A, A^c, \Omega\}$이며 $\{1_B \in D\} \in \{\emptyset, B, B^c, \Omega\}$이므로 16가지를 확인하면 충분하다. 수반된 집합이 $\emptyset$ 또는 Ω이면 자명하게 등식이 성립하며, 남은 경우는 (i)에 의해 등식이 성립한다. □

첫째 정의가 둘째 정의의 특수한 경우이며 둘째 정의가 셋째 정의의 특수한 경우이므로 여러 대상들이 독립이라는 것이 무엇을 의미하는지를 정의할 때는 반대 순서로 정의할 것이다. 먼저 유한 개 대상들의 경우로 문제를 줄이는 것으로 시작하자. 대상(σ-체, 확률변수, 집합)들의 무한 족이 독립이라는 것의 정의는 모든 유한 부분족이 독립인 것이다.

σ-체 $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n$이 **독립(independent)**이라는 것의 정의는 $i = 1, \ldots, n$에 대하여 $A_i \in \mathcal{F}_i$인 경우 다음이 성립하는 것이다.

$$P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$$

확률변수 $X_1, \ldots, X_n$이 **독립(independent)**이라는 것의 정의는 $i = 1, \ldots, n$에 대하여 $B_i \in \mathcal{R}$인 경우 다음이 성립하는 것이다.

$$P\left(\bigcap_{i=1}^{n}\{X_i \in B_i\}\right) = \prod_{i=1}^{n} P(X_i \in B_i)$$

집합 $A_1, \ldots, A_n$이 **독립(independent)**이라는 것의 정의는 $I \subset \{1, \ldots, n\}$이면 다음이 성립하는 것이다.

$$P\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} P(A_i)$$

언뜻 보기에는 마지막 정의가 다른 두 가지와 맞지 않는 듯하다. 그러나 잠시 생각해 보면 지시확률변수 $1_{A_i}, 1 \le i \le n$들이 독립이라 하고 $i \in I$이면 $B_i = \{1\}$, $i \notin I$이면 $B_i = \mathbb{R}$로 취할 경우 위 정의의 조건이 따라온다는 사실을 알 수 있다. 역으로 다음이 성립한다:

Theorem 2.1.3. $A_1, A_2, \ldots, A_n$이 독립이라 하자.
(i) $A_1^c, A_2, \ldots, A_n$이 독립이다.
(ii) $1_{A_1}, \ldots, 1_{A_n}$이 독립이다.

Proof. $B_1 = A_1^c$이며 $i > 1$에 대하여 $B_i = A_i$라 하자. 만약 $I \subset \{1, \ldots, n\}$이 1을 포함하지 않으면 $P(\bigcap_{i\in I} B_i) = \prod_{i\in I} P(B_i)$임은 자명하다. 이제 $1 \in I$이며 $J = I - \{1\}$이라 하자. $P(\bigcap_{i\in J} A_i) = \prod_{i\in J} P(A_i)$에서 $P(\bigcap_{i\in I} A_i)$를 빼면 $P(A_1^c \cap \bigcap_{i\in J} A_i) = P(A_1^c)\prod_{i\in J} P(A_i)$를 얻는다.
(ii) (i)을 반복적용하면 만약 $B_i \in \{A_i, A_i^c\}$인 경우 $B_1, \ldots, B_n$이 독립임을 알 수 있다. 그러므로 만약 $C_i \in \{A_i, A_i^c, \Omega\}$이면 $P(\bigcap_{i=1}^n C_i) = \prod_{i=1}^n P(C_i)$이다. 어떠한 $C_i = \emptyset$이면 마지막 등식이 자명하게 성립하므로 $1_{A_i} \in \{\emptyset, A_i, A_i^c, \Omega\}$라는 사실에 의해 결과가 따라온다. □

독립 사건의 정의에서 이해해야 할 점 중 한 가지는 모든 $i \neq j$에 대하여 $P(A_i \cap A_j) = P(A_i)P(A_j)$라 가정하는 것으로는 충분하지 않다는 점이다. 이러한 성질을 가지는 사건들의 열 $A_1, \ldots, A_n$은 **쌍마다 독립(pairwise independent)**이라 불린다. 독립 사건들이 쌍마다 동립임은 자명하다. 다음 예시는 역이 참이 아님을 보여준다.

Example 2.1.4. X_1, X_2, X_3이 독립 확률변수이며 다음을 만족시킨다 하자.

$$P(X_i = 0) = P(X_i = 1) = 1/2$$

$A_1 = \{X_2 = X_3\}, A_2 = \{X_3 = X_1\}, A_3 = \{X_1 = X_2\}$라 하자. 만약 $i \neq j$이면 다음이 성립하므로 이러한 사건들은 쌍마다 독립이다.

$$P(A_i \cap A_j) = P(X_1 = X_2 = X_3) = 1/4 = P(A_i)P(A_j)$$

그러나 다음이 성립하므로 이들은 독립이 아니다.

$$P(A_1 \cap A_2 \cap A_3) = 1/4 \neq 1/8 = P(A_1)P(A_2)P(A_3)$$

확률변수 X, Y가 독립임을 확인하기 위해서는 $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$임을 모든 Borel 집합 A, B에 대하여 확인해야 한다. Borel 집합은 지나치게 많으므로 우리의 다음 주제는 이를 개선하는 것이다:

2.1.1 Sufficient Conditions for Independence (독립성의 충분조건)

우리의 주요 결과는 Theorem 2.1.7이다. 이 결과를 기술하기 위해서는 앞선 정의들을 일반화하는 정의가 필요하다.

집합족 $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n \subseteq \mathcal{F}$가 **독립(independent)**이라는 것의 정의는 $A_i \in \mathcal{A}_i$이며 $I \subset \{1, \ldots, n\}$이면 $P(\bigcap_{i\in I} A_i) = \prod_{i\in I} P(A_i)$인 것이다.

만약 각각의 족이 하나의 집합으로 구성되면, i.e. $\mathcal{A}_i = \{A_i\}$이면 이러한 정의는 집합에 대한 독립성의 정의가 된다.

Lemma 2.1.5. 일반성을 잃지 않고 각각의 $\mathcal{A}_i$가 Ω를 포함한다 가정할 수 있다. 이 경우 (각각의 $i \notin I$에 대하여 $A_i = \Omega$로 설정할 수 있으므로) 위 조건은 다음과 동치이다.

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

Proof. $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$이 독립이며 $\bar{\mathcal{A}}_i = \mathcal{A}_i \cup \{\Omega\}$라 하자. 만약 $A_i \in \bar{\mathcal{A}}_i$이며 $I = \{j : A_j \neq \Omega\}$이면 $\bigcap_i A_i = \bigcap_{i \in I} A_i$이므로 $\bar{\mathcal{A}}_1, \bar{\mathcal{A}}_2, \ldots, \bar{\mathcal{A}}_n$도 독립이다. □

Theorem 2.1.7의 증명은 Dynkin π-λ 정리에 기반한다. 그 결과를 기술하기 위해서는 두 가지 정의가 필요하다. $\mathcal{A}$가 **π-계(π-system)**라는 것은 교집합 하에서 닫혀 있는 것이다. i.e. 만약 $A, B \in \mathcal{A}$이면 $A \cap B \in \mathcal{A}$이다. $\mathcal{L}$이 **λ-계(λ-system)**라는 것은 다음을 만족시키는 것이다: (i) $\Omega \in \mathcal{L}$ (ii) 만약 $A, B \in \mathcal{L}$이며 $A \subset B$이면 $B - A \in \mathcal{L}$ (iii) 만약 $A_n \in \mathcal{L}$이며 $A_n \uparrow A$이면 $A \in \mathcal{L}$이다.

Theorem 2.1.6. 만약 $\mathcal{P}$가 π-계이며 $\mathcal{L}$이 $\mathcal{P}$를 포함하는 λ-계이면 $\sigma(\mathcal{P}) \subset \mathcal{L}$이다.

증명은 Appendix의 Section A.1에 숨겨 놓겠다.

Theorem 2.1.7. $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$이 독립이며 각각의 $\mathcal{A}_i$가 π-계라 하자. 그 경우 $\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2), \ldots, \sigma(\mathcal{A}_n)$이 독립이다.

Proof. $A_2, \ldots, A_n$이 $A_i \in \mathcal{A}_i$를 만족시키는 집합들이며 $F = A_2 \cap \cdots \cap A_n$이고 $\mathcal{L} = \{A : P(A \cap F) = P(A)P(F)\}$라 하자. $P(\Omega \cap F) = P(\Omega)P(F)$이므로 $\Omega \in \mathcal{L}$이다. λ-계의 정의의 (ii)를 확인하기 위해 만약 $A, B \in \mathcal{L}$이며 $A \subset B$이면 $(B - A) \cap F = (B \cap F) - (A \cap F)$임을 기억해 두라. 그러므로 Theorem 1.1.1(i)을 적용한 후 $A, B \in \mathcal{L}$에 대하여 이를 한 번 더 적용하면 다음을 얻는다.

$$\begin{aligned} P((B - A) \cap F) &= P(B \cap F) - P(A \cap F) = P(B)P(F) - P(A)P(F) \\ &= \{P(B) - P(A)\}P(F) = P(B - A)P(F) \end{aligned}$$

따라서 $B - A \in \mathcal{L}$이다. (iii)을 확인하기 위해 $B_k \in \mathcal{L}$이며 $B_k \uparrow B$라 가정하자. $(B_k \cap F) \uparrow (B \cap F)$이므로 Theorem 1.1.1(iii)을 적용한 후 $B_k \in \mathcal{L}$에 대하여 이를 한 번 더 적용하면 다음을 얻는다.

$$P(B \cap F) = \lim_k P(B_k \cap F) = \lim_k P(B_k)P(F) = P(B)P(F)$$

이제 $\pm$-λ 정리를 적용하면 $\mathcal{L} \supset \sigma(\mathcal{A}_1)$이다. $A_1 \in \sigma(\mathcal{A}_1)$이며 $A_i \in \mathcal{A}_i$ $(2 \leq i \leq n)$이면 다음이 성립함이 따라온다.

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)P\left(\bigcap_{i=2}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

이제 Lemma 2.15를 사용하면 다음을 얻는다:

(*) 만약 $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$이 독립이면 $\sigma(\mathcal{A}_1), \mathcal{A}_2, \ldots, \mathcal{A}_n$이 독립이다.

(*)를 $\mathcal{A}_2, \ldots, \mathcal{A}_n, \sigma(\mathcal{A}_1)$에 적용하면 (독립성의 정의는 순서를 치환해도 변하지 않으므로 이는 독립이다) $\sigma(\mathcal{A}_2), \mathcal{A}_3, \ldots, \mathcal{A}_n, \sigma(\mathcal{A}_2)$가 독립임을 보여준다. n회 반복하면 요구된 결과를 얻는다. □

Remark. 독자는 $A, B \in \mathcal{L}$인 경우 $A \cap B \in \mathcal{L}$ 또는 $A \cup B \in \mathcal{L}$임을 보이는 것은 쉽지 않은 반면에 $A, B \in \mathcal{L}, A \subset B$이면 $B - A \in \mathcal{L}$임을 확인하는 것은 간단함을 기억해야 한다.

Theorem 2.1.7의 증명 과정으로부터 여러 따름정리들을 얻는다.

Theorem 2.1.8. $X_1, \ldots, X_n$이 독립임을 확인하기 위해서는 모든 $x_1, \ldots, x_n \in (-\infty, \infty]$에 대하여 다음을 확인하면 충분하다.

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{i=1}^{n} P(X_i \leq x_i)$$

Proof. $\mathcal{A}_i$가 $\{X_i \leq x_i\}$ 형태의 집합들의 족이라 하자. $(x \wedge y)_i = x_i \wedge y_i = \min\{x_i, y_i\}$라 하면 다음이 성립하므로,

$$\{X_i \leq x\} \cap \{Y_i \leq y\} = \{X_i \leq x \cap y\}$$

$\mathcal{A}_i$는 π-계이다. $x_i = \infty$인 경우를 허용했으므로 $\Omega \in \mathcal{A}_i$이다. Exercise 1.3.1은 $\sigma(\mathcal{A}_i) = \sigma(X_i)$임을 함의하며 따라서 결과가 Theorem 2.1.7에서 따라온다. □

이러한 결과는 확률변수의 독립성을 이들의 확률분포함수를 통해 표현한다. Exercise 2.1.1과 2.1.2는 확률밀도함수들과 이산 확률변수들을 다룬다.

우리의 다음 목표는 독립 확률변수들의 서로 소 족에 대한 함수들이 독립임을 증명하는 것이다. 엄밀한 진술을 위해서는 Theorem 2.1.10을 참조하라. 먼저 우리는 σ-체에 대한 유사한 결과를 증명할 것이다.

Theorem 2.1.9. $\mathcal{F}_{i,j}$ $(1 \le i \le n, 1 \le j \le m(i))$들이 독립이며 $\mathcal{G}_i = \sigma(\bigcup_j \mathcal{F}_{i,j})$라 하자. 그 경우 $\mathcal{G}_1, \dots, \mathcal{G}_n$이 독립이다.

Proof. $\mathcal{A}_i$가 $\bigcap_j A_{i,j}$ $(A_{i,j} \in \mathcal{F}_{i,j})$ 형태의 집합들의 족이라 하자. $\mathcal{A}_i$는 π-계이며 Ω를 포함하고 $\bigcup_j \mathcal{F}_{i,j}$를 포함한다. 따라서 Theorem 2.1.7은 $\sigma(\mathcal{A}_i) = \mathcal{G}_i$들이 독립임을 함의한다. □

Theorem 2.1.10. $1 \le i \le n, 1 \le j \le m(i)$에 대하여 $X_{i,j}$들이 독립이며 $f_i : \mathbb{R}^{m(i)} \to \mathbb{R}$들이 가측 함수라 하자. 그 경우 $f_i(X_{i,1}, \dots, X_{i,m(i)})$들이 독립이다.

Proof. $\mathcal{F}_{i,j} = \sigma(X_{i,j}), \mathcal{G}_i = \sigma(\bigcup_j \mathcal{F}_{i,j})$라 하자. $f_i(X_{i,1}, \dots, X_{i,m(i)}) \in \mathcal{G}_i$이므로 요구된 결과는 Theorem 2.1.9와 Exercise 2.1.1에서 따라온다. □

다음과 같은 Theorem 2.1.10의 구체적인 특수한 경우를 곧 사용할 것이다: 만약 $X_1, \dots, X_n$이 독립이면 $X = X_1$과 $Y = X_2, \dots, X_n$이 독립이다. 나중에 우리는 독립 확률변수 $X_1, \dots, X_n$의 합 $S_m = X_1 + \cdots + X_m$들을 다루면서 Theorem 2.1.10을 활용하여 만약 $m < n$이면 $S_n - S_m$이 사건 $\{\max_{1 \le k \le m} S_k > x\}$의 지시함수와 독립이라 결론지을 것이다.

2.1.2 Independence, Distribution, and Expectation (독립성, 확률분포, 기댓값)

우리의 다음 목표는 독립 확률변수들의 확률분포와 기댓값에 대한 공식을 얻는 것이다.

Theorem 2.1.11. $X_1, \dots, X_n$이 독립 확률변수이며 X_i가 확률분포 μ_i를 가지면 $(X_1, \dots, X_n)$이 분포함수 $\mu_1 \times \cdots \times \mu_n$을 가진다.

Proof. (i) $A_1 \times \cdots \times A_n$, (ii) 독립성, (iii) μ_i, (iv) $\mu_1 \times \cdots \times \mu_n$의 정의를 사용하면 다음을 얻는다.

$$\begin{aligned} P((X_1, \dots, X_n) \in A_1 \times \cdots \times A_n) &= P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i) \\ &= \prod_{i=1}^n \mu_i(A_i) = \mu_1 \times \cdots \times \mu_n(A_1 \times \cdots \times A_n) \end{aligned}$$

마지막 공식은 $(X_1, \dots, X_n)$의 확률분포와 측도 $\mu_1 \times \cdots \times \mu_n$이 $A_1 \times \cdots \times A_n$ 형태의 집합들에서 일치함을 보여주며, 이러한 집합들의 족은 $\mathcal{R}^n$을 생성하는 π-계이다. 따라서 Theorem 2.1.6에 의해 이들이 일치해야 한다. □

Theorem 2.1.12. X, Y가 독립이며 확률분포 μ, ν를 가진다 하자. 만약 $h : \mathbb{R}^2 \to \mathbb{R}$이 가측 함수이며 $h \ge 0$ 또는 $E|h(X,Y)| < \infty$를 만족시킬 경우 다음이 성립한다.

$$Eh(X,Y) = \iint h(x,y)\, \mu(dx)\, \nu(dy)$$

특히 만약 $f, g \ge 0$ 또는 $E|f(X)|, E|g(Y)| < \infty$를 만족시키는 가측함수 $f, g : \mathbb{R} \to \mathbb{R}$에 대하여 $h(x,y) = f(x)g(y)$일 경우 다음이 성립한다.

$$Ef(X)g(Y) = Ef(X) \cdot Eg(Y)$$

Proof. Theorem 1.6.9와 Fubini 정리(Theorem 1.7.2)를 사용하면 다음을 얻는다.

$$Eh(X,Y) = \int_{\mathbb{R}^2} h\, d(\mu \times \nu) = \iint h(x,y)\, \mu(dx)\, \nu(dy)$$

둘째 결과를 증명하기 위해 $f, g \ge 0$일 때의 결과로 시작하겠다. 이 경우 첫째 결과와 $g(y)$가 x에 의존하지 않는다는 사실을 사용하고 Theorem 1.6.9를 2회 적용하면 다음을 얻는다.

$$\begin{aligned} Ef(X)g(Y) &= \iint f(x)g(y)\, \mu(dx)\, \nu(dy) = \int g(y) \int f(x)\, \mu(dx)\, \nu(dy) \\ &= \int Ef(X)g(y)\, \nu(dy) = Ef(X)Eg(Y) \end{aligned}$$

음 아닌 f, g에 대한 결과를 $|f|, |g|$에 적용하면 $E|f(X)g(Y)| = E|f(X)|E|g(Y)| < \infty$임을 보여준다. 우리는 마지막 논의를 반복하여 요구된 결과를 증명할 수 있다. $\square$

Theorem 2.1.12에서 조금만 더 나가면 다음을 얻는다:

Theorem 2.1.13. 만약 $X_1, \ldots, X_n$이 독립이며 (a) 모든 i에 대하여 $X_i \geq 0$ 또는 (b) 모든 i에 대하여 $E|X_i| < \infty$라 하자. 그 경우 다음이 성립한다.

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} EX_i$$

i.e. 좌변의 기댓값이 존재하며 우변과 같은 값을 가진다.

Proof. Theorem 2.1.10에 의해 $X = X_1$과 $Y = X_2 \cdots X_n$이 독립이다. 따라서 $f(x) = |x|, g(y) = |y|$로 선택하면 $E|X_1 \cdots X_n| = E|X_1|E|X_2 \cdots X_n|$이다. 귀납법에 의해 $1 \leq m \leq n$에 대하여 다음이 성립함이 따라온다.

$$E|X_m \cdots X_n| = \prod_{i=m}^{n} E|X_k|$$

만약 $X_i \geq 0$이면 $|X_i| = X_i$이며 요구된 결과가 특수한 경우 $m = 1$에서 따라온다. 일반적인 결과를 증명하기 위해, $m = 2$인 특수한 경우가 $E|Y| = E|X_2 \cdots X_n| < \infty$를 보여줌을 기억해 두라. 그러므로 $f(x) = x, g(y) = y$에 대하여 Theorem 2.1.12를 사용하면 $E(X_1 \cdots X_n) = EX_1 \cdot E(X_2 \cdots X_n)$을 얻으며 요구된 결과는 귀납법에서 따라온다. $\square$

Example 2.1.14. 독립이 아닌 확률변수에 대하여 $E(XY) = EX \cdot EY$일 수 있다. X와 Y의 결합 확률분포가 다음 표에 의해 주어졌다 하자.

		Y		
		1	0	-1
	1	0	a	0
X	0	b	c	b
	-1	0	a	0

여기에서 $a, b > 0, c \geq 0, 2a + 2b + c = 1$이다. 이들은 $XY \equiv 0$이도록 조정되었다. 대칭성에 의해 $EX = 0, EY = 0$이므로 $E(XY) = 0 = EXEY$이다. 그러나 다음에 의해 이들 확률변수는 독립이 아니다.

$$P(X = 1, Y = 1) = 0 < ab = P(X = 1)P(Y = 1)$$

$EX^2, EY^2 < \infty$이며 $EXY = EXEY$를 만족시키는 두 확률변수 X, Y를 **비상관(uncorrelated)**이라 한다. 2차 모멘트가 유한하다는 조건은 Cauchy-Schwarz 부등식을 통해 $E|XY| < \infty$임을 보장하기 위해 필요하다.

2.1.3 Sums of Independent Random Variables (독립 확률변수들의 합)

Theorem 2.1.15. 만약 X, Y가 독립이며 $F(x) = P(X \leq x), G(y) = P(Y \leq y)$이면 다음이 성립한다.

$$P(X + Y \leq z) = \int F(z - y)\, dG(y)$$

우변의 적분은 F와 G의 **합성곱(convolution)**이라 불리며 $F * G(z)$로 표기된다. $dG(y)$의 의미는 증명에서 설명하겠다.

Proof. $h(x, y) = 1_{(x+y \leq z)}$라 하자. μ, ν가 확률분포함수 F, G를 가지는 확률측도라 하자. 고정된 y에 대하여 다음이 성립하므로,

$$\int h(x, y)\, \mu(dx) = \int_{(-\infty, z-y]}(x)\, \mu(dx) = F(z - y)$$

Theorem 2.1.12를 사용하면 다음을 얻는다.

$$\begin{aligned} P(X + Y \leq z) &= \iint 1_{(x+y \leq z)}\, \mu(dx)\, \nu(dy) \\ &= \int F(z - y)\, \nu(dy) = \int F(z - y)\, dG(y) \end{aligned}$$

마지막 등식은 표기법 변경에 불과하다: 우리는 $dG(y)$를 '확률분포함수 G를 가지는 측도 ν에 대한 적분'의 축약으로 간주한다. □

구체적인 예시를 다루기 위해서는 Theorem 2.1.15의 특수한 경우가 필요하다.

Theorem 2.1.16. X가 확률밀도함수 f를 가지며 Y가 확률분포함수 G를 가지고 이들이 독립이라 하자. 그 경우 $X+Y$는 다음의 확률밀도함수를 가진다.

$$h(x) = \int f(x-y)\, dG(y)$$

만약 Y가 확률밀도함수 g를 가지는 경우 위 공식은 다음과 같이 표현될 수 있다.

$$h(x) = \int f(x-y)g(y)\, dy$$

Proof. Theorem 2.1.15, 확률밀도함수의 정의, (모든 것이 음 아닌 함수이므로) Fubini 정리(Theorem 1.7.2)에 의해 다음이 성립한다.

$$\begin{aligned} P(X+Y \le z) &= \int F(z-y)\, dG(y) = \int \int_{-\infty}^{z} f(x-y)\, dx\, dG(y) \\ &= \int_{-\infty}^{z} \int f(x-y)\, dG(y)\, dx \end{aligned}$$

마지막 등식은 $X+Y$가 확률밀도함수 $h(x) = \int f(x-y)\, dG(y)$를 가짐을 보여준다. 둘째 공식은 Theorem 2.1.15의 증명에서 주어진 $dG(y)$의 의미를 상기하고 Exercise 1.6.8을 사용하면 따라온다. □

Theorem 2.1.16에 더해 지저분한 계산을 하는 것으로 두 가지 표준적인 예시를 다룰 수 있다. 이러한 사실들은 학부 확률론에서 배워서 익숙할 것이다.

Example 2.1.17. 매개변수 α, λ를 가지는 **감마 확률밀도함수(gamma density)**는 다음에 의해 주어진다.

$$f(x) = \begin{cases} \lambda^\alpha x^{\alpha-1} e^{-\lambda x}/\Gamma(\alpha) & (x \ge 0) \\ 0 & (x < 0) \end{cases}$$

여기에서 $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$이다.

Theorem 2.1.18. 만약 $X = \text{gamma}(\alpha,\lambda), Y = \text{gamma}(\beta,\lambda)$이 독립이면 $X+Y = \text{gamma}(\alpha+\beta,\lambda)$이다. 결과적으로 만약 $X_1,\dots,X_n$이 독립 exponential(λ) 확률변수들이면 $X_1+\cdots+X_n$은 gamma(n,λ) 확률분포를 가진다.

Proof. $X+Y$의 확률밀도함수를 $f_{X+Y}(z)$로 표기하고 Theorem 2.1.16을 사용하면 다음을 얻는다.

$$\begin{aligned} f_{X+Y}(x) &= \int_0^x \frac{\lambda^\alpha (x-y)^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda(x-y)} \frac{\lambda^\beta y^{\beta-1}}{\Gamma(\beta)} e^{-\lambda y}\, dy \\ &= \frac{\lambda^{\alpha+\beta} e^{-\lambda x}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x (x-y)^{\alpha-1} y^{\beta-1}\, dy \end{aligned}$$

그러므로 적분이 $x^{\alpha+\beta+1}\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$임을 보이면 충분하다. 이를 보이기 위해 $y = xu, dy = x\, du$로 변수변환하여 다음을 얻는 것으로 시작하겠다.

$$x^{\alpha+\beta-1} \int_0^1 (1-u)^{\alpha-1} u^{\beta-1}\, du = \int_0^x (x-y)^{\alpha-1} y^{\beta-1}\, dy \tag{2.1.1}$$

이 지점에서 증명을 완료할 수 있는 방법은 두 가지가 있다. 간단한 방법은 확률밀도함수가 다음과 같은 상수 $c_{\alpha,\beta}$에 대하여 $f_{X+Y}(x) = c_{\alpha,\beta} e^{-\lambda x} \lambda^{\alpha+\beta} x^{\alpha+\beta-1}$임을 보였음을 기억해 두는 것이다.

$$c_{\alpha,\beta} = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 (1-u)^{\alpha-1} u^{\beta-1}\, du$$

f_{X+Y}를 확률분포함수로 만드는 정규화 상수는 $c_{\alpha,\beta}$ 하나뿐이므로 베타 분포의 정의를 상기하면 $c_{\alpha,\beta} = 1/\Gamma(\alpha+\beta)$여야 한다.

베타 분포의 정의를 기억해내지 못하는 독자를 위한 덜 명쾌한 방법은 위 등식을 미적분학을 통해 증명하는 것이다. 위 방정식의 양변에 e^{-x}를 곱하고 0에서 ∞까지 적분한 후 우변에 Fubini 정리를 사용하면 다음을 얻는다.

$$\begin{aligned}\Gamma(\alpha+\beta)\int_0^1 (1-u)^{\alpha-1}u^{\beta-1}\,du &= \int_0^\infty \int_0^x y^{\beta-1}e^{-y}(x-y)^{\alpha-1}e^{-(x-y)}\,dy\,dx \\ &= \int_0^\infty y^{\beta-1}e^{-y}\int_y^\infty (x-y)^{\alpha-1}e^{-(x-y)}\,dx\,dy = \Gamma(\alpha)\Gamma(\beta)\end{aligned}$$

이는 첫째 결과를 제공한다. 둘째 결과는 gamma$(1,\lambda)$가 exponential(λ)라는 사실과 귀납법에서 따라온다. □

Example 2.1.19 (Normal distribution (정규분포))**.** Example 1.6.11에서 평균 μ와 분산 a를 가지는 정규확률밀도함수를 도입했다.

$$(2\pi a)^{-1/2}\exp(-(x-\mu)^2/2a)$$

Theorem 2.1.20. 만약 $X = \text{normal}(\mu, a), Y = \text{normal}(\nu, b)$가 독립이면 $X + Y = \text{normal}(\mu+\nu, a+b)$이다.

Proof. $\mu = \nu = 0$인 경우에 대하여 결과를 증명하면 충분하다. $Y_1 = \text{normal}(0,a), Y_2 = \text{normal}(0,b)$라 하자. 그 경우 Theorem 2.1.16은 다음을 함의한다.

$$f_{Y_1+Y_2}(z) = \frac{1}{2\pi\sqrt{ab}}\int e^{-x^2/2a}e^{-(z-x)^2/2b}\,dx$$

앞의 상수를 제외하면 적분은 다음과 같이 재표현될 수 있다.

$$\begin{aligned}&\int \exp\left(-\frac{bx^2+ax^2-2axz+az^2}{2ab}\right)\,dx \\ &\quad= \int \exp\left(-\frac{a+b}{2ab}\left\{x^2 - \frac{2a}{a+b}xz + \frac{a}{a+b}z^2\right\}\right)\,dx \\ &\quad= \int \exp\left(-\frac{a+b}{2ab}\left\{\left(x-\frac{a}{a+b}z\right)x^2 + \frac{ab}{(a+b)^2}z^2\right\}\right)\,dx\end{aligned}$$

(여기에서 $-\{a/(a+b)\}^2 + \{a/(a+b)\} = ab/(a+b)^2$가 사용되었다.) x에 의존하지 않는 항을 적분 밖으로 내보내면 위 적분은 다음과 같아진다.

$$\begin{aligned}&= \exp\left(-\frac{z^2}{2(a+b)}\right)\int \exp\left(-\frac{a+b}{2ab}\left(x-\frac{a}{a+b}z\right)^2\right)\,dx \\ &= \exp\left(-\frac{z^2}{2(a+b)}\right)\sqrt{2\pi ab/(a+b)}\end{aligned}$$

마지막 적분은 정규화 상수를 제외한 매개변수 $\mu = az/(a+b), \sigma^2 = ab/(a+b)$를 가지는 정규확률밀도함수의 적분이다. 처음에 제외했던 상수를 재도입하면 다음을 얻는다.

$$f_{Y_1+Y_2}(z) = \frac{1}{2\pi\sqrt{ab}}\sqrt{2\pi ab/(a+b)}\exp\left(-\frac{z^2}{2(a+b)}\right)$$

□

2.1.4 Constructing Independent Random Variables (독립 확률변수의 구축)

독립 확률변수를 연구하기 전에 다뤄야 할 마지막 문제는 다음과 같다: 이들이 존재하는가? (만약 이들이 존재하지 않는다면 이들을 연구해야 할 이유가 없다!) 만약 유한 개 분포함수 F_i $(1 \le i \le n)$들이 주어졌다면

$P(X_i \leq x) = F_i(x)$를 만족시키는 독립 확률변수 $X_1, \ldots, X_n$을 간단히 구축할 수 있다. $\Omega = \mathbb{R}^n, \mathcal{F} = \mathcal{R}^n, X_i(\omega_1, \ldots, \omega_n) = \omega_i$ ($\omega \in \mathbb{R}^n$의 i번째 좌표)이며 P가 다음과 같은 $\mathcal{R}^n$에서의 측도라 하자.

$$P((a_1, b_1] \times \cdots \times (a_n, b_n]) = (F_1(b_1) - F_1(a_1)) \cdots (F_n(b_n) - F_n(a_n))$$

μ_i가 확률분포함수 F_i를 가지는 측도라 하면 $P = \mu_1 \times \cdots \times \mu_n$이다.

우리는 주어진 확률분포함수들을 가지는 독립 확률변수들의 무한 열 $X_1, X_2, \ldots$를 구축하기 위해 위 구축을 다음의 무한 곱공간에서 수행하고자 한다.

$$\mathbb{R}^{\mathbb{N}} = \{(\omega_1, \omega_2, \ldots) : \omega_i \in \mathbb{R}\} = \{\text{함수 } \omega : \mathbb{N} \to \mathbb{R}\}$$

여기에서 $\mathbb{N} = \{1, 2, \cdots\}$는 **자연수(natural number)** 집합이다. $X_i(\omega) = \omega_i$로 정의하고 $\mathbb{R}^{\mathbb{N}}$에 **유한 차원 집합(finite dimensional set)**(i.e. $B_i \in \mathcal{R}$에 대하여 $\{\omega : \omega_i \in B_i, 1 \leq i \leq n\}$ 형태의 집합)들에 의해 생성된 곱 σ-체 $\mathcal{R}^{\mathbb{N}}$을 부여할 것이다. $\mathcal{R}^{\mathbb{N}}$으로의 유일한 확장의 존재성을 보장하기 위해서는 Theorem A.3.1이 필요하다:

Theorem 2.1.21 (Kolmogorov's extension theorem (Kolmogorov 확장 정리))**.** $(\mathbb{R}^n, \mathcal{R}^n)$에서의 확률측도 μ_n들이 일관적으로 주어졌다 하자. i.e.

$$\mu_{n+1}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R}) = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n])$$

그 경우 $(\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}})$에서의 유일한 확률측도 P가 존재하여 다음을 만족시킨다.

$$P(\omega : \omega_i \in (a_i, b_i], 1 \leq i \leq n) = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n])$$

다음으로 우리는 다른 가측 공간 $(S, \mathcal{S})$에 속한 값을 가지는 확률변수들의 열을 구축해야 한다. 불운하게도 Theorem 2.1.21은 임의의 가측 공간에 대해서는 성립하지 않는다. (서로 다른 공간들의 무한 곱 $\Omega_1 \times \Omega_2 \times \cdots$에서의) 첫 번째 예시는 Andersen and Jessen (1948)에 의해 구축되었다. (Halmos (1950) p.214 또는 Neveu (1965) p. 84를 참조하라.) 공간 Ω_i들이 모두 같은 경우를 위해서는 Wegner (1973)을 참조하라. 다행히도 우리의 모든 결과에 대하여 적합하며 Kolmogorov 정리가 자명하게 일반화되도록 하는 공간들의 족이 존재한다.

$(S, \mathcal{S})$가 **우량(nice)**임은 S에서 $\mathbb{R}$로의 단사 함수 φ가 존재하여 φ와 φ^{-1}이 모두 가측이도록 하는 것이다.

이러한 공간들은 때로는 **통상 Borel 공간(standard Borel space)**이라고도 불리지만, 이미 Borel의 이름을 딴 것이 너무 많다. 다음 결과는 응용에서 등장하는 거의 모든 공간들이 우량임을 보여준다.

Theorem 2.1.22. 만약 S가 완비 가분 거리공간 M의 Borel 부분집합이며 $\mathcal{S}$가 S의 Borel 부분집합들의 족이면 $(S, \mathcal{S})$가 우량이다.

Proof. 다음 거리함수를 가지는 특수한 경우 $S = [0, 1)^{\mathbb{N}}$으로 시작하자.

$$\rho(x, y) = \sum_{n=1}^{\infty} |x_n - y_n| / 2^n$$

만약 $x = (x^1, x^2, x^3, \ldots)$이면 각각의 성분을 이진 소수 $x^j = .x_1^j x_2^j x_3^j \ldots$로 전개하자. (무한히 많은 0을 가지는 전개를 취하자.) 다음과 같다 하자.

$$\varphi_o(x) = .x_1^1 x_2^1 x_1^2 x_3^1 x_2^2 x_1^3 x_4^1 x_3^2 x_2^3 x_1^4 \ldots$$

일반적인 경우를 다루기 위해 다음과 같다 하면,

$$d(x, y) = \rho(x, y) / (1 + \rho(x, y))$$

(세부사항을 위해서는 Exercise 2.1.3을 참조하라) 모든 x, y에 대하여 거리 $d(x, y) < 1$이라 가정할 수 있다. $q_1, q_2, \ldots$가 Sd ㅔ서의 가산 조밀 집합이라 하자. 다음과 같다 하자.

$$\psi(x) = (d(x, q_1), d(x, q_2), \ldots)$$

$\psi : S \to [0, 1)^{\mathbb{N}}$이 단사 연속 함수이다. $\varphi_o \circ \psi$가 요구된 함수를 제공한다. □

Caveat emptor (경고). 이곳에서의 증명은 세부사항들이 생략되어 있다. 더 포괄적인 논의를 위해서는 Dudley (1989)의 section 13.1을 참조하라. 이곳에서의 해석학의 흥미로운 결과는 연속체 가설이 참이도록 하는 완비 가분 거리공간의 Borel 부분집합들이 존재한다는 것이다: i.e. 모든 집합이 유한집합, 가산 무한집합, 또는 실수의 기수를 가진다.

Exercises (연습문제)

2.2 Weak Laws of Large Numbers (대수의 약법칙)

이 절에서 우리는 여러 가지 '대수의 약법칙'들을 증명할 것이다. 첫 번째 과제는 정리의 결론에서 등장할 수렴의 방식을 정의하는 것이다. Y_n이 Y로 **측도에서 수렴(converge in probability)**한다는 것의 정의는 모든 $\varepsilon > 0$에 대하여 $n \to \infty$에서 $P(|Y_n - Y| > \varepsilon) \to 0$인 것이다.

2.2.1 L^2 Weak Laws (L^2 대수의 약법칙)

우리는 분산을 계산하고 Chebyshev 부등식을 사용하여 얻어진 대수의 약법칙들로 시작하겠다. Example 2.1.14에서 2개의 확률변수에 대하여 주어진 정의를 확장하여, $EX_i^2 < \infty$를 만족시키는 확률변수들의 족 X_i $(i \in I)$가 **비상관(uncorrelated)**이라는 것의 정의는 다음이 성립하는 것이다.

$$E(X_iX_j) = EX_iEX_j \qquad (i \neq j)$$

비상관 확률변수에 대한 대수의 약법칙(Theorem 2.2.3)의 핵심은 다음과 같다:

Theorem 2.2.1. $X_1, \ldots, X_n$이 $E(X_i^2) < \infty$를 만족시키며 비상관이라 하자. 그 경우 다음이 성립한다.

$$\text{var}(X_1 + \cdots + X_n) = \text{var}(X_1) + \cdots + \text{var}(X_n)$$

여기에서 $\text{var}(Y)$는 Y의 분산을 의미한다.

Proof. $\mu_i = EX_i$이며 $S_n = \sum_{i=1}^n X_i$라 하자. $ES_n = \sum_{i=1}^n \mu_i$이므로 분산의 정의를 사용하고 합의 제곱을 합의 2개 사본의 곱으로 표현한 후 전개하면 다음을 얻는다.

$$\begin{aligned}
\text{var}(S_n) &= E(S_n - ES_n)^2 = E\left(\sum_{i=1}^n (X_i - \mu_i)\right)^2 \\
&= E\left(\sum_{i=1}^n \sum_{j=1}^n (X_i - \mu_i)(X_j - \mu_j)\right) \\
&= \sum_{i=1}^n E(X_i - \mu_i)^2 + 2\sum_{i=1}^n \sum_{j=1}^{i-1} E((X_i - \mu_i)(X_j - \mu_j))
\end{aligned}$$

여기에서 마지막 등식은 대각 항 $i = j$를 분리한 후 $1 \le i < j \le n$에 대한 합이 $1 \le j < i \le n$에 대한 합과 같다는 사실을 이용한 것이다.

첫째 합은 $\text{var}(X_1) + \cdots + \text{var}(X_n)$이며 따라서 우리는 둘째 합이 0임을 보이고자 한다. 이를 수행하기 위해서는 X_i와 X_j과 비상관이므로 다음이 성립함을 관찰하라.

$$\begin{aligned}
E((X_i - \mu_i)(X_j - \mu_j)) &= EX_iX_j - \mu_iEX_j - \mu_jEX_i + \mu_i\mu_j \\
&= EX_iX_j - \mu_i\mu_j = 0
\end{aligned}$$

□

말로 설명하면, Theorem 2.2.1은 비상관 확률변수들에 대하여 합의 분산이 분산의 합임을 말해 준다. Theorem 2.2.3에 대한 우리의 증명의 두 번째 구성 요소는 다음과 같은 (1.6.4)의 결과이다:

$$\text{var}(cY) = c^2 \,\text{var}(Y)$$

세 번째 구성 요소는 다음과 같다.

Lemma 2.2.2. 만약 $p > 0$이며 $E|Z_n|^p \to 0$이면 $Z_n \to 0$으로 확률에서 수렴한다.

Proof. Chebyshev 부등식(Theorem 1.6.4)을 $\varphi(x) = x^p$와 $X = |Z_n|$에 대하여 적용하면 $\varepsilon > 0$에 대하여 $P(|Z_n| \ge \varepsilon) \le \varepsilon^{-p} E|Z_n|^p \to 0$임을 보여준다. □

Theorem 2.2.3 (L^2 대수의 약법칙)**.** $X_1, X_2, \ldots$가 비상관 확률변수이며 $EX_i = \mu$이고 $\text{var}(X_i) \le C < \infty$라 하자. 만약 $S_n = X_1 + \cdots + X_n$이면 $n \to \infty$에서 $S_n/n \to \mu$로 L^2에서 수렴하며 확률에서 수렴한다.

Proof. L^2 수렴성을 증명하기 위해 $E(S_n/n) = \mu$이며 따라서 다음이 성립함을 관찰하라.

$$E(S_n/n - \mu)^2 = \mathrm{var}(S_n/n) = \frac{1}{n^2}(\mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n)) \le \frac{Cn}{n^2} \to 0$$

확률에서의 수렴성도 성립한다 결론짓기 위해 $Z_n = S_n/n - \mu$에 대하여 Lemma 2.2.2를 적용하라. □

Theorem 2.2.3의 가장 중요한 특수한 경우는 $X_1, X_2, \ldots$가 모두 동일한 확률분포함수를 가지는 독립 확률변수들일 때 발상한다. 이 경우 확률변수들이 **독립동분포(independent and identically distributed)** 또는 간단히 **i.i.d.**라 한다. Theorem 2.2.3은 이 경우에 $EX_i^2 < \infty$이면 $n \to \infty$에서 S_n/n이 $\mu = EX_i$로 확률에서 수렴함을 보여준다. Theorem 2.2.14에서 우리는 결론을 얻기 위해서는 $E|X_i| < \infty$이면 충분함을 보일 것이지만 잠시 동안 더 약한 결과를 다루겠다.

첫 번째 응용은 표면적으로는 무작위성이 드러나지 않는 상황에 대한 것이다.

3 | Central Limit Theorems (중심극한정리)

4 | Martingales (마팅게일)

마팅게일 X_n은 공정한 게임에 돈을 거는 참가자의 시간 n에서의 운으로 간주될 수 있다; 열마팅게일(우마팅게일)은 유리한(불리한) 게임의 결과물로 간주될 수 있다. 마팅게일에 대한 두 가지 기본적인 사실들이 있다. 첫째는 여기에 베팅하는 것으로 돈을 벌 수 없다는 것이다. (Theorem 4.2.8을 참조하라.) 특히 만약 참가자가 어떠한 유계 시간 N 내에 게임을 중단하기로 선택했을 경우 승리 기댓값 EX_N은 초기 운 X_0와 같다. (여기에서 잠시 동안 X_0가 무작위가 아니라 가정하는 중이다.) 두 번째 사실인 Theorem 4.2.11은 열마팅게일에 대한 것이다. Mike Brennan에 의한 휴리스틱을 사용하면 "이들은 비감소 열의 확률적인 유사품이며 따라서 이들이 위로 유계이면 (엄밀하게는 $\sup_n EX_n^+ < \infty$이면) 이들은 거의 확실하게 수렴한다." Section 4.3의 내용이 보여주는 바와 같이 이러한 결과는 여러 응용을 가진다. 뒤쪽 절들에서 우리는 마팅게일이 L^p $(p > 1)$(Section 4.4)와 L^1(Section 4.6)에서 수렴할 충분조건을 제시하고 제곱 적분가능 마팅게일의 특수한 경우를 연구하며 (Section 4.5) $n \le 0$에 의해 첨자화된 마팅게일을 고려할 것이다. (Section 4.7) 우리는 비유계 정지시간에 대하여 $EX_N = EX_0$가 성립할 충분조건을 제시할 것이다. (Section 4.8) 이러한 결과들은 난보의 거동을 연구하는 데 상당히 유용하다. Section 4.9는 조합론적 증명을 제시하는 것으로 Section 4.8.1에서의 마팅게일 논의로부터 유도된 난보 결과들을 보충할 것이다.

4.1 Conditional Expectation (조건부기댓값)

이 장과 다음 장에서 중요한 정의로 시작하겠다. 정의를 제시한 이후에는 이를 설명하는 여러 예시들을 고려할 것이다. 확률공간 $(\Omega, \mathcal{F}_0, P)$, σ-체 $\mathcal{F} \subset \mathcal{F}_0$, $E|X| < \infty$인 확률변수 $X \in \mathcal{F}_0$가 주어졌다 하자. **X의 $\mathcal{F}$에서의 조건부기댓값(conditional expectation of X given $\mathcal{F}$)** $E(X|\mathcal{F})$는 다음을 만족시키는 임의의 확률변수 Y이다:

(i) $Y \in \mathcal{F}$이다. i.e. $\mathcal{F}$가 가측이다.
(ii) 모든 $A \in \mathcal{F}$에 대하여 $\int_A X\, dP = \int_A Y\, dP$이다.

(i)과 (ii)를 만족시키는 임의의 Y는 **$E(X|\mathcal{F})$의 일종(version of $E(X|\mathcal{F})$)**이라 불린다. 가장 먼저 다뤄야 할 문제는 조건부기댓값의 존재성과 유일성이다. 우리는 유일성을 먼저 공략할 것이지만 기술적인 문제로 시작하겠다.

Lemma 4.1.1. 만약 Y가 (i), (ii)를 만족시키면 이는 적분가능하다.

Proof. $A = \{Y > 0\} \in \mathcal{F}$라 하고 (ii)를 2회 적용한 후 다음을 더하자.

$$\int Y\, dP = \int_A X\, dP \le \int_A |X|\, dP$$
$$\int_{A^c} -Y\, dP = \int_{A^c} -X\, dP \le \int_{A^c} |X|\, dP$$

따라서 $E|Y| \le E|X|$이다. □

Uniqueness (유일성). 만약 Y'도 (i), (ii)를 만족시킨다면 다음이 성립한다.

$$\int_A Y\, dP = \int_A Y'\, dP \qquad (\forall A \in \mathcal{F})$$

$A = \{Y - Y' \ge \varepsilon > 0\}$으로 선택하면 다음을 얻는다.

$$0 = \int_A X - X\, dP = \int_A Y - Y'\, dP \ge \varepsilon P(A)$$

따라서 $P(A) = 0$이다. 이것이 모든 ε에 대하여 성립하므로 $Y \leq Y'$ a.s.이다. Y와 Y'의 역할을 교환하면 $Y = Y'$ a.s.를 얻는다. 기술적으로 $Y = E(X|\mathcal{F})$와 같은 모든 등식들은 $Y = E(X|\mathcal{F})$ a.s.로 표기되어야 하나 우리는 이전 장들에서 이를 무시했으며 앞으로도 그렇게 할 것이다.

마지막 논의를 반복하면 다음을 얻는다:

Theorem 4.1.2. 만약 $B \in \mathcal{F}$에서 $X_1 = X_2$이면 $E(X_1|\mathcal{F}) = E(X_2|\mathcal{F})$ a.s.B이다.

Proof. $Y_1 = E(X_1|\mathcal{F}), Y_2 = E(X_2|\mathcal{F})$라 하자. $A = \{Y_1 - Y_2 \geq \varepsilon > 0\}$으로 취하면 다음을 얻는다.

$$0 = \int_{A\cap B} X_1 - X_2 \, dP = \int_{A\cap B} Y_1 - Y_2 \, dP \geq \varepsilon P(A)$$

따라서 $P(A) = 0$이며 앞에서와 같이 결론이 따라온다. □

Existence (존재성). 먼저 ν가 μ**에 대하여 절대연속(absolutely continuous with respect to** μ**)** 이라는 것은 ($\nu \ll \mu$로 축약되며) $\mu(A) = 0$이면 $\nu(A) = 0$임을 의미함을 상기하라. 우리는 Theorem A.4.8 을 사용할 것이다:

Radon-Nikodym Theorem (Radon-Nikodym 정리). μ, ν가 $(\Omega, \mathcal{F})$에서의 σ-유한 측도라 하자. 만약 $\nu \ll \mu$이면 함수 $f \in \mathcal{F}$가 존재하여 모든 $A \in \mathcal{F}$에 대하여 다음을 만족시킨다.

$$\int_A f \, d\mu = \nu(A)$$

f는 통상적으로 $d\nu/d\mu$로 표기되며 **Radon-Nikodym 도함수(Radon-Nikodym derivative)**라 불린다.

위 정리는 조건부기댓값의 존재성을 간단히 보장한다. 먼저 $X \geq 0$이라 하자. $\mu = P$이며 다음과 같다 하자.

$$\nu(A) = \int_A X \, dP \qquad (\forall A \in \mathcal{F})$$

지배 수렴 정리는 ν가 측도임을 함의하며 (Exercise 1.5.4를 참조하라) 적분의 정의는 $\nu \ll \mu$를 함의한다. Radon-Nikodym 도함수 $d\nu/d\mu \in \mathcal{F}$는 임의의 $A \in \mathcal{F}$에 대하여 다음을 만족시킨다.

$$\int_A X \, dP = \nu(A) = \int_A \frac{d\nu}{d\mu} \, dP$$

$A = \Omega$로 선택하면 $d\nu/d\mu \geq 0$이 적분가능함을 알 수 있으며 우리는 $d\nu/d\mu$가 $E(X|\mathcal{F})$의 일종임을 보였다.

이제 일반적인 경우를 다루자. $X = X^+ - X^-$이며 $Y_1 = E(X^+|\mathcal{F}), Y_2 = E(X^-|\mathcal{F})$라 하자. $_2 - Y_2 \in \mathcal{F}$ 는 적분가능하며 임의의 $A \in \mathcal{F}$에 대하여 다음을 만족시킨다.

$$\begin{aligned}\int_A X \, dP &= \int_A X^+ \, dP - \int_A X^- \, dP \\ &= \int_A Y_1 \, dP - \int_A Y_2 \, dP = \int_A (Y_1 - Y_2) \, dP\end{aligned}$$

이는 $Y_1 - Y_2$가 $E(X|\mathcal{F})$의 일종임을 보여주며 증명을 완료한다. □

4.1.1 Examples (예시)

직관적으로 우리는 $\mathcal{F}$가 우리가 처리할 수 있는 정보를 기술하는 것으로 생각할 수 있다 - 각각의 $A \in \mathcal{F}$에 대하여 우리는 A가 발생하거나 그렇지 않음을 안다. 그 경우 $E(X|\mathcal{F})$는 우리가 가진 정보 하에서 X의 값에 대한 '최적 추측'이다. 몇 가지 예시들은 이를 명확히 하고 $E(X|\mathcal{F})$를 조건부기댓값의 다른 정의와 관련짓는 것을 도울 것이다.

Example 4.1.3. 만약 $X \in \mathcal{F}$이면 $E(X|\mathcal{F}) = X$이다; i.e. X를 알면 '최적 추측'은 X 자신이다. X가 항상 (ii)를 만족시키므로 X가 $E(X|\mathcal{F})$가 아니도록 할 수 있는 유일한 장애물은 조건 (i)이다. 이러한 예시의 특수한 경우는 상수확률변수 $X = c$이다.

Example 4.1.4. 완전한 정보의 반대쪽 극단은 정보가 없는 것이다. X가 $\mathcal{F}$에 독립이라 하자. i.e. 모든 $B \in \mathcal{R}$과 $A \in \mathcal{F}$에 대하여 다음이 성립한다.

$$P(\{X \in B\} \cap A) = P(X \in B)P(A)$$

우리는 이 경우 $E(X|\mathcal{F}) = EX$임을 주장하겠다; i.e. X에 대하여 아무것도 모르면 최적 추측은 평균 EX이다. 정의를 확인하면 $EX \in \mathcal{F}$이며 따라서 (i)이 성립함을 기억해 두라. (ii)를 검증하기 위해 만약 $A \in \mathcal{F}$이면 X와 $1_A \in \mathcal{F}$가 독립이므로 Theorem 2.1.13에 의해 다음이 성립함을 기억해 두라.

$$\int_A X\, dP = E(X1_A) = EXE1_A = \int_A EX\, dP$$

독자는 이곳에서와 앞으로의 내용에서 게임은 '추측 및 검증'임을 기억해 두어야 한다. 우리는 조건부기댓값에 대한 공식을 찾아내고 이것이 (i)과 (ii)를 만족시킴을 확인할 것이다.

Example 4.1.5. 이 예시에서 우리는 조건부기댓값의 새로운 정의를 학부 확률론에서 배운 고전적인 정의와 관련지을 것이다. $\Omega_1, \Omega_2, \ldots$가 Ω의 서로 소 집합들로의 유한 또는 무한 분할이며 각각 양수 확률을 가지고 $\mathcal{F} = \sigma(\Omega_1, \Omega_2, \ldots)$가 이러한 집합들에 의해 생성된 σ-체라 하자. 그 경우 Ω_i에서 다음이 성립한다.

$$E(X|\mathcal{F}) = \frac{E(X;\Omega_i)}{P(\Omega_i)}$$

말로 설명하면, Ω_i에 대한 정보는 결과물이 어느 분할에 속하는지를 말해 주며 이러한 정보 하에서 X의 최적 추측은 X의 Ω_i 상에서의 평균임을 말해 준다. 이러한 추측이 맞음을 증명하기 위해 제시된 공식이 각각의 Ω_i에서 상수이며 따라서 이것이 $\mathcal{F}$에 대하여 가측임을 관찰하라. (ii)를 검증하기 위해서는 $A = \Omega$인 경우에 등식을 확인하면 충분하다. 그러나 이는 자명하다:

$$\int_{\Omega_i} \frac{E(X;\Omega_i)}{P(\Omega_i)}\, dP = E(X;\Omega_i) = \int_{\Omega_i} X\, dP$$

퇴화되었지만 중요한 특수한 경우는 자명 σ-체 $\mathcal{F} = \{\emptyset, \Omega\}$이다. 이 경우 $E(X|\mathcal{F}) = EX$이다.

학부 개념과의 연관성을 계속 다루자. 다음과 같다 하자.

$$P(A|\mathcal{G}) = E(1_A|\mathcal{G})$$
$$P(A|B) = P(A \cap B)/P(B)$$

위 예시에서 Ω_i에서 $P(A|\mathcal{F}) = P(A|\Omega_i)$였음을 관찰하라.

주어진 σ-체에서의 조건부 확률의 특수한 경우로 확률변수에 의한 통제가 있다. 이는 다음과 같이 정의된다.

$$E(X|Y) = E(X|\sigma(Y))$$

여기에서 $\sigma(Y)$는 Y에 의해 생성된 σ-체이다.

Example 4.1.6. 학부 확률론에서의 조건부기댓값의 정의와의 관련성을 수립하는 작업을 계속하여, X와 Y가 결합확률밀도함수 $f(x,y)$를 가진다 하자. i.e.

$$P((X,Y) \in B) = \int_B f(x,y)\, dx\, dy \qquad (B \in \mathcal{R}^2)$$

간결성을 위해 모든 y에 대하여 $\int f(x,y)\, dx > 0$이라 하자. 이 경우 만약 $E|g(X)| < \infty$이며 h가 다음에 의해 정의된다면 $E(g(X)|Y) = h(Y)$임을 주장하겠다.

$$h(y) = \int g(x)f(x,y)\, dx \bigg/ \int f(x,y)\, dx$$

이 공식을 '추측'하기 위해서는 확률밀도함수 $P(Y = y)$를 실제 확률처럼 다루면 된다.

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f(x,y)}{\int f(x,y)\, dx}$$

그러므로 조건부확률밀도함수에 대하여 적분하면 다음을 얻는다.

$$E(g(X)|Y=y) = \int g(x)P(X=x|Y=y)\,dx$$

이제 제시된 공식을 '검증'하기 위하여 $h(Y) \in \sigma(Y)$이며 (i)이 성립함을 기억해 두라. (ii)를 확인하기 위해 $A \in \sigma(Y)$이면 어떠한 $B \in \mathcal{R}$에 대하여 $A = \{\omega : Y(\omega) \in B\}$이며 따라서 다음이 성립함을 관찰하라.

$$\begin{aligned} E(h(X);A) &= \int_B \int h(y)f(x,y)\,dx\,dy = \int_B \int g(x)f(x,y)\,dx\,dy \\ &= E(g(X)1_B(Y)) = E(g(X);A) \end{aligned}$$

Remark. $\int f(x,y)\,dx > 0$이라는 전제조건을 제거하기 위해서는 h를 다음과 같이 정의하라.

$$h(y)\int f(x,y)\,dx = \int g(y)f(x,y)\,dx$$

(i.e. $\int f(x,y) = 0$인 곳에서는 h가 무엇이든 될 수 있다.) 그리고 이 정의에 대하여 위 증명을 적용 가능함을 관찰하라.

Example 4.1.7. X와 Y가 독립이라 하자. φ가 $E|\varphi(X,Y)| < \infty$를 만족시키는 함수이며 $g(X) = E(\varphi(x,Y))$라 하자. 우리는 이제 다음이 성립함을 보일 것이다.

$$E(\varphi(X,Y)|X) = g(X)$$

Proof. $g(X) \in \sigma(X)$임은 명백하다. (ii)를 확인하기 위해 만약 $A \in \sigma(X)$이면 $A = \{X \in C\}$임을 기억해 두라. 따라서 변수변환 공식(Theorem 1.6.9)와 (X,Y)의 결합확률분포가 곱측도라는 사실(Theorem 2.1.11)을 사용하고 g의 정의를 사용한 후 다시 변수변환을 적용하면 다음을 얻는다.

$$\begin{aligned} \int_A \varphi(X,Y)\,dP &= E\{\varphi(X,Y)1_C(X)\} \\ &= \iint \varphi(x,y)1_C(x)\,\nu(dx)\,\mu(dy) \\ &= \int 1_C(x)g(x)\,\mu(dx) = \int_A g(X)\,dP \end{aligned}$$

이는 요구된 결과를 증명한다. □

Example 4.1.8 (Borel's paradox (Borel 역설))**.** X가 지구상의 무작위로 선택된 점이며 θ가 그 경도, φ가 그 위도라 하자. 관습적으로 $\theta \in [0,2\pi), \varphi \in (-\pi/2,\pi/2]$이도록 선택한다. 그러나 우리는 이와 동치로 $\theta \in [0,\pi)$이며 $\varphi \in (-\pi,\pi]$이도록 선택할 수도 있다. 다른 말로 하면 새로운 경도는 점이 속한 대원을 명시하며 그 후 φ가 각도를 제공한다.

언뜻 보기에는 X가 지구 상에서 균일하다면 θ와 대원 상에서의 각도 φ가 모두 가능한 값들에 대하여 균일해야 할 것처럼 보인다. θ는 균일하지만 φ는 그렇지 않다. 두 가지 방법 모두에서 φ가 θ에 독립적이며 따라서 조건부확률분포가 비조건부확률분포와 같지만 북극보다 적도 근처에 더 땅이 많으므로 이것이 불균일함을 알아차린다면 이러한 역설이 완전히 사라진다.

4.1.2 Properties (성질)

조건부기댓값은 통상적인 기댓값과 동일한 여러 성질들을 가진다.

Theorem 4.1.9. (a)와 (b)에서 $E|X|, E|Y| < \infty$라 가정하겠다.
(a) 조건부 확률이 선형적이다:

$$E(aX+Y|\mathcal{F}) = aE(X|\mathcal{F}) + E(Y|\mathcal{F}) \tag{4.1.1}$$

(b) 만약 $X \le Y$이면 다음이 성립한다.

$$E(X|\mathcal{F}) \le E(Y|\mathcal{F}) \tag{4.1.2}$$

(c) 만약 $X_n \ge 0$이며 $X_n \uparrow X$이고 $EX < \infty$이면 다음이 성립한다.

$$E(X_n|\mathcal{F}) \uparrow E(X|F) \tag{4.1.3}$$

Remark. 마지막 결과를 $Y_1 - Y_n$에 적용하면 $Y_n \downarrow Y$이며 $E|Y_1|, E|Y| < \infty$인 경우 $E(Y_n|\mathcal{F}) \downarrow E(Y|\mathcal{F})$임을 알 수 있다.

Proof. (a)를 증명하기 위해서는 우변이 좌변의 일종임을 확인해야 한다. 이는 명백히 $\mathcal{F}$-가측이다. (ii)를 확인하기 위해 만약 $A \in \mathcal{F}$이면 적분의 선형성과 $E(X|\mathcal{F}), E(Y|\mathcal{F})$를 정의하는 성질에 의해 다음이 성립함을 관찰하라.

$$\begin{aligned}\int_A \{aE(X|\mathcal{F}) + E(Y|\mathcal{F})\}\, dP &= a\int_A E(X|\mathcal{F})\, dP + \int_A E(Y|\mathcal{F})\, dP \\ &= \int_A X\, dP + \int_A Y\, dP = \int_A aX + Y\, dP\end{aligned}$$

이는 (4.1.1)을 증명한다.

다음 정의를 사용하자.

$$\int_A E(X|\mathcal{F})\, dP = \int_A X\, dP \le \int_A Y\, dP = \int_A E(Y|\mathcal{F})\, dP$$

$A = \{E(X|\mathcal{F}) - E(Y|\mathcal{F}) \ge \varepsilon > 0\}$이라 하면 모든 $\varepsilon > 0$에 대하여 이러한 집합이 확률 0을 가짐을 알 수 있고 따라서 (4.1.2)가 증명된다.

$Y_n = X - X_n$이라 하자. $E(Y_n|\mathcal{F}) \downarrow 0$임을 보이면 충분하다. $Y_n \downarrow$이므로 (4.1.2)는 $Z_n \equiv E(Y_n|\mathcal{F})$가 어떠한 극한 $Z_n \downarrow Z_\infty$로 수렴함을 함의한다. 만약 $A \in \mathcal{F}$이면 다음이 성립한다.

$$\int_A Z_n\, dP = \int_A Y_n\, dP$$

$n \to \infty$라 하고 $Y_n \downarrow 0$임을 유의하여 지배 수렴 정리를 사용하면 모든 $A \in \mathcal{F}$에 대하여 $\int_A Z_\infty\, dP = 0$이며 따라서 $Z_\infty \equiv 0$임을 보여준다. □

Theorem 4.1.10. 만약 φ가 볼록이며 $E|X|, E|\varphi(X)| < \infty$이면 다음이 성립한다.

$$\varphi(E(X|\mathcal{F})) \le E(\varphi(X)|\mathcal{F}) \tag{4.1.4}$$

Proof. 만약 φ가 선형이면 결과는 자명하며 따라서 우리는 φ가 비선형이라 가정할 것이다. 이러한 가정하에서 $S = \{(a,b) : a, b \in \mathbb{Q}, \forall x\ ax + b \le \varphi(x)\}$라 하면 $\varphi(x) = \sup\{ax + b : (a,b) \in S\}$이다. 세부사항을 위해서는 Theorem 1.6.2의 증명을 참조하라. 만약 $\varphi(x) \ge ax + b$이면 (4.1.2)와 (4.1.1)은 다음을 함의한다.

$$E(\varphi(X)|\mathcal{F}) \ge aE(X|\mathcal{F}) + b \qquad \text{a.s.}$$

$(a,b) \in S$ 상에서의 상한을 취하면 다음을 얻는다.

$$E(\varphi(X)|\mathcal{F}) \ge \varphi(E(X|\mathcal{F})) \qquad \text{a.s.}$$

이는 요구된 결과를 제공한다. □

Remark. 여기에서 우리는 각각의 a, b에 대한 예외적인 경우가 존재하며 따라서 가산집합 상에서 상한을 취해야 함을 강조하기 위해 부등식에 a.s.를 표기했다.

Theorem 4.1.11. 조건부기댓값은 $L^p, p \ge 1$에서의 수축이다.

Proof. (4.1.4)는 $|E(X|\mathcal{F})|^p \le E(|X|^p|\mathcal{F})$임을 함의한다. 기댓값을 취하면 다음을 얻는다.

$$E(|E(X|\mathcal{F})|^p) \le E(E(|X|^p|\mathcal{F})) = E|X|^p$$

□

마지막 부등식에서 우리는 정의에서 즉시 따라오는 다음의 항등식을 사용했다. (정의의 성질 (ii)를 $A = \Omega$에 대하여 적용하라.)

$$E(E(Y|\mathcal{F})) = E(Y) \tag{4.1.5}$$

조건부기댓값은 (4.1.5)와 같이 '통상적인' 기댓값에 대한 유사품이 존재하지 않는 성질들도 가진다.

Theorem 4.1.12. 만약 $\mathcal{F} \subset \mathcal{G}$이며 $E(X|\mathcal{G}) \in \mathcal{F}$이면 $E(X|\mathcal{F}) = E(X|\mathcal{G})$이다.

Proof. 가정에 의해 $E(X|\mathcal{G}) \in \mathcal{F}$이다. 정의의 다른 부분을 확인하기 위해서는 $A \in \mathcal{F} \subset \mathcal{G}$에 대하여 다음이 성립함을 기억해 두면 충분하다.

$$\int_A X\, dP = \int_A E(X|\mathcal{G})\, dP$$

□

Theorem 4.1.13. 만약 $\mathcal{F}_1 \subset \mathcal{F}_2$이면 (i) $E(E(X|\mathcal{F}_1)|\mathcal{F}_2) = E(X|\mathcal{F}_1)$이다.
(ii) $E(E(X|\mathcal{F}_2)|\mathcal{F}_1) = E(X|\mathcal{F}_1)$이다.

다른 말로 하면 더 작은 σ-체가 항상 이긴다. 증명에서 보게 될 것과 같이 첫째 등식은 자명하다. 둘째 등식은 간단히 증명할 수 있지만 Theorem 4.1.14와 조합하면 조건부기댓값을 계산하는 강력한 도구가 된다. 이는 거짓인 결과들을 증명하기 위해 빈번히 사용된다.

Proof. $E(X|\mathcal{F}_1) \in \mathcal{F}_2$이므로 (i)은 Example 4.1.3에서 따라온다. (ii)를 증명하기 위해서는 $E(X|\mathcal{F}_1) \in \mathcal{F}_1$임과 $A \in \mathcal{F}_1 \subset \mathcal{F}_2$이면 다음이 성립함을 기억해 두라.

$$\int_A E(X|\mathcal{F}_1)\, dP = \int_A X\, dP = \int_A E(X|\mathcal{F}_2)\, dP$$

□

다음 결과는 $\mathcal{F}$에 대한 조건부기댓값에 대하여 확률변수 $X \in \mathcal{F}$들이 상수함수처럼 거동함을 보여준다; 이들을 '적분' 밖으로 가져올 수 있다.

Theorem 4.1.14. 만약 $X \in \mathcal{F}$이며 $E|Y|, E|XY| < \infty$이면 다음이 성립한다.

$$E(XY|\mathcal{F}) = XE(Y|\mathcal{F})$$

Proof. 우변은 $\in \mathcal{F}$이므로 (ii)를 확인해야 한다. 이를 위해 통상적인 4단계 과정을 사용할 것이다. 먼저 $X = 1_B$ $(B \in \mathcal{F})$라 하자. 이 경우 만약 $A \in \mathcal{F}$이면 다음이 성립한다.

$$\int_A 1_B E(Y|\mathcal{F})\, dP = \int_{A\cap B} E(Y|\mathcal{F})\, dP = \int_{A\cap B} Y\, dP = \int_A 1_B Y\, dP$$

따라서 (ii)가 성립한다. 선형성에 의해 이 결과를 단순확률변수 X로 확장할 수 있다. 만약 $X, Y \geq 0$이면 $X_n \uparrow X$가 X로 수렴하는 단순확률변수열이라 하고 단조 수렴 정리를 사용하여 다음이 성립한다 결론지을 수 있다.

$$\int_A XE(Y|\mathcal{F})\, dP = \int_A XY\, dP$$

일반적인 결과를 증명하기 위해서는 X와 Y를 양수부와 음수부로 분해하면 된다. □

Theorem 4.1.15. $EX^2 < \infty$라 가정하자. $E(X|\mathcal{F})$는 '평균제곱오차' $E(X-Y)^2$를 최소화하는 확률변수 $Y \in \mathcal{F}$이다.

Remark. 이 결과는 $E(X|\mathcal{F})$의 '기하학적 해석'을 제공한다. $L^2(\mathcal{F}_o) = \{Y \in \mathcal{F}_o : EY^2 < \infty\}$는 Hilbert 공간이며 $L^2(\mathcal{F})$는 닫힌 부분공간이다. 이 경우 $E(X|\mathcal{F})$는 X의 $L^2(\mathcal{F})$로의 사영이다. 즉 X에 가장 가까운 부분공간의 점이다.

Proof. 먼저 만약 $Z \in L^2(\mathcal{F})$이면 Theorem 4.1.14가 다음을 함의함을 관찰하자.

$$ZE(X|\mathcal{F}) = E(ZX|\mathcal{F})$$

(Cauchy-Schwarz 부등식에 의해 $E|XZ| < \infty$이다.) 기댓값을 취하면 다음을 얻는다.

$$E(ZE(X|\mathcal{F})) = E(E(ZX|\mathcal{F})) = E(ZX)$$

정리하면 다음과 같다.

$$E[Z(X - E(X|\mathcal{F}))] = 0 \qquad (Z \in L^2(\mathcal{F}))$$

만약 $Y \in L^2(\mathcal{F})$이며 $Z = E(X|\mathcal{F}) - Y$이면 교차항이 사라지므로 다음이 성립한다.

$$E(X-Y)^2 = E\{X - E(X|\mathcal{F}) + Z\}^2 = E\{X - E(X|\mathcal{F})\}^2 + EZ^2$$

마지막 공식으로부터 $Z = 0$일 경우 $E(X-Y)^2$가 최소화됨을 간단히 보일 수 있다. □

4.1.3 Regular Conditional Probabilities (정칙조건부확률)*

생략

Exercises (연습문제)

4.2 Martingales, Almost Sure Convergence (마팅게일, 거의 확실한 수렴)

이 절에서 우리는 마팅게일과 그 유사품인 우마팅게일 및 열마팅게일을 정의하고 이들에 관한 이론을 개발하기 시작할 것이다. $\mathcal{F}_n$이 **여과(filtration)**, 즉 σ-체들의 증가 열이라 하자. 열 X_n이 $\mathcal{F}_n$에 **적응(adapted)**했음의 정의는 모든 n에 대하여 $X_n \in \mathcal{F}_n$인 것이다. 만약 X_n이 다음을 만족시키는 열이면,

(i) $E|X_n| < \infty$
(ii) X_n이 $\mathcal{F}_n$에 적응했다.
(iii) 모든 n에 대하여 $E(X_{n+1}|\mathcal{F}_n) = X_n$이다.

X가 ($\mathcal{F}_n$에 대한) **마팅게일(martingale)**이라 한다. 만약 마지막 정의에서 =가 $\leq$ 또는 $\geq$로 대체된 경우 각각 X가 **우마팅게일(supermartingale)** 또는 **열마팅게일(submartingale)**이라 한다.

난보와 관련된 세 가지 예시를 기술하는 것으로 시작하겠다. $\xi_1, \xi_2, \ldots$가 독립동분포라 하자. S_0이 상수이며 $S_n = S_0 + \xi_1 + \cdots + \xi_n$이라 하자. $n \geq 1$에 대하여 $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$이라 하고 $\mathcal{F}_0 = \{\emptyset, \Omega\}$로 취하자.

Example 4.2.1 (Linear martingale (선형마팅게일))**.** 만약 $\mu = E\xi_i = 0$이면 $n \geq 0$에 대한 S_n들은 $\mathcal{F}_n$에 대한 마팅게일이다.

이를 증명하기 위해 $S_n \in \mathcal{F}_n, E|S_n| < \infty$이며 ξ_{n+1}이 $\mathcal{F}_n$에 독립임을 관찰하라. 따라서 조건부기댓값의 선형성(4.1.1)과 Example 4.1.4를 이용하면 다음을 얻는다.

$$E(S_{n+1}|\mathcal{F}_n) = E(S_n|\mathcal{F}_n) + E(\xi_{n+1}|\mathcal{F}_n) = S_n + E\xi_{n+1} = S_n$$

만약 $\mu \leq 0$이면 위 계산은 $E(X_{n+1}|\mathcal{F}_n) \leq X_n$, i.e. X_n이 우마팅게일임을 보여준다. 이 경우 X_n은 불리한 게임에 돈을 거는 것에 대응하며 따라서 우마팅게일에서는 '우월'한 점이 없다. 이러한 명칭은 만약 f가 우조화함수(superharmonic, i.e. f가 2계 이하의 연속 도함수들을 가지며 $\partial^2 f/\partial x_1^2 + \cdots + \partial^2 f/\partial x_d^2 \leq 0$)이면 다음이 성립한다는 사실에서 나온다.

$$f(x) \geq \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)\, dy$$

여기에서 $B(x,r) = \{y : |x-y| < r\}$은 반지름 r의 공이며 $|B(x,r)|$은 이러한 공의 부피이다.

만약 $\mu \geq 0$이면 S_n은 열마팅게일이다. 첫째 결과를 $\xi_i' = \xi_i - \mu$에 적용하면 $S_n - n\mu$가 마팅게일임을 알 수 있다.

Example 4.2.2 (Quadratic martingale (2차마팅게일))**.** 이제 $\mu = E\xi_i = 0$이며 $\sigma^2 = \text{var}(\xi_i) < \infty$라 하자. 이 경우 $S_n^2 - n\sigma^2$가 마팅게일이다.

$(S_n + \xi_{n+1})^2 = S_n^2 + 2S_n\xi_{n+1} + \xi_{n+1}^2$이며 ξ_{n+1}이 $\mathcal{F}_n$과 독립이므로 다음이 성립한다.

$$\begin{aligned} E(S_{n+1}^2 - (n+1)\sigma^2|\mathcal{F}_n) &= S_n^2 + 2S_nE(\xi_{n+1}|\mathcal{F}_n) + E(\xi_{n+1}^2|\mathcal{F}_n) - (n+1)\sigma^2 \\ &= S_n^2 + 0 + \sigma^2 - (n+1)\sigma^2 = S_n^2 - n\sigma^2 \end{aligned}$$

Example 4.2.3 (Exponential martingale (지수마팅게일))**.** $Y_1, Y_2, \ldots$가 음 아닌 i.i.d. 확률변수들이며 $EY_m = 1$이라 하자. 만약 $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$이면 $M_n = \prod_{m \leq n} Y_m$은 마팅게일을 정의한다. 이를 증명하기 위해서는 다음이 성립함을 기억해 두라.

$$E(M_{n+1}|\mathcal{F}_n) = M_n E(X_{n+1}|\mathcal{F}_n) = Y_n$$

이제 $Y_i = e^{\theta\xi_i}$이며 $\phi(\theta) = Ee^{\theta\xi_i} < \infty$라 하자. $Y_i = \exp(\theta\xi)/\phi(\theta)$가 평균 1을 가지므로 $EY_i = 1$이고 따라서 다음이 성립한다.

$$M_n = \prod_{i=1}^n Y_i = \exp(\theta S_n)/\phi(\theta)^n \quad \text{이 마팅게일이다.}$$

우리는 아래에서 다른 많은 예시들을 보게 될 것이며 따라서 우리는 이제 마팅게일의 성질들을 유도할 것이다. 우리의 첫째 결과는 우마팅게일의 정의에서 즉시 따라온다. 우리는 이 결과를 우마팅게일의 정의로 취할 수도 있지만, 그 경우 정의에 부합하는지를 확인하는 것이 힘들어질 것이다.

Theorem 4.2.4. 만약 X_n이 우마팅게일이면 $n > m$에 대하여 $E(X_n|\mathcal{F}_m) \le X_m$이다.

Proof. 정의는 $n = m + 1$인 경우에 대한 결과를 제공한다. $n = m + k$이며 $k \ge 2$라 하자. Theorem 4.1.2, 정의, (4.1.2)에 의해 다음이 성립한다.

$$E(X_{m+k}|\mathcal{F}_m) = E(E(X_{m+k}|\mathcal{F}_{m+k-1})|\mathcal{F}_m) \le E(X_{m+k-1}|\mathcal{F}_m)$$

요구된 결과는 귀납법에 의해 따라온다. □

Theorem 4.2.5. (i) 만약 X_n이 열마팅게일이면 $n > m$에 대하여 $E(X_n|\mathcal{F}_m) \ge X_m$이다.
(ii) 만약 X_n이 마팅게일이면 $n > m$에 대하여 $E(X_n|\mathcal{F}_m) = X_m$이다.

Proof. (i)을 증명하기 위해서는 $-X_n$이 우마팅게일임을 기억해 두고 (4.1.1)을 사용하라. (ii)를 위해서는 X_n이 우마팅게일이며 열마팅게일임을 기억해 두라. □

Remark. Theorem 4.2.5의 증명의 발상은 앞으로 여러 번 사용될 것이다. 불필요한 반복을 피하기 위해 우리는 결과를 우마팅게일 또는 열마팅게일에 대해서만 기술한 후 다른 쪽으로 번역하는 것은 독자에게 남길 것이다.

Theorem 4.2.6. 만약 X_n이 $\mathcal{F}_n$에 대한 마팅게일이며 φ가 볼록함수이고 모든 n에 대하여 $E|\varphi(X_n)| < \infty$이면 $\varphi(X_n)$이 $\mathcal{F}_n$에 대한 열마팅게일이다. 결과적으로 만약 $p \ge 1$이며 모든 n에 대하여 $E|X_n|^p < \infty$이면 $|X_n|^p$는 $\mathcal{F}_n$에 대한 열마팅게일이다.

Proof. Jensen 부등식과 정의에 의해 다음이 성립한다.

$$E(\varphi(X_{n+1})|\mathcal{F}_n) \ge \varphi(E(X_{n+1}|\mathcal{F}_n)) = \varphi(X_n)$$

□

Theorem 4.2.7. 만약 X_n이 $\mathcal{F}_n$에 대한 열마팅게일이며 φ가 모든 n에 대하여 $E|\varphi(X_n)| < \infty$를 만족시키는 증가 볼록함수이면 $\varphi(X_n)$이 $\mathcal{F}_n$에 대한 열마팅게일이다. 따라서 (i) 만약 X_n이 열마팅게일이면 $(X_n - a)^+$가 열마팅게일이다. (ii) 만약 X_n이 우마팅게일이면 $X_n \wedge a$가 우마팅게일이다.

Proof. Jensen 부등식과 가정에 의해 다음이 성립한다.

$$E(\varphi(X_{n+1})|\mathcal{F}_n) \ge \varphi(E(X_{n+1}|\mathcal{F}_n)) \ge \varphi(X_n)$$

□

$\mathcal{F}_n$ $(n \ge 0)$이 여과라 하자. H_n $(n \ge 1)$이 **예측 가능 열(predictable seuqence)**이라는 것의 정의는 모든 $n \ge 1$에 대하여 $H_n \in \mathcal{F}_{n-1}$인 것이다. 다른 말로 하면 H_n의 값은 시행 $n - 1$에서의 정보에 의해 (확실히) 결정된다. 이 절에서 우리는 H_n을 도박사가 시간 n에 거는 돈의 양으로 간주할 것이다. 이는 시간 $1, \ldots, n-1$의 결과에 기반하며 시간 n의 결과에 기반하지 않는다!

H_n을 도박 체계로 간주하면 이러한 체계 하에서 돈을 얼마나 벌 수 있는지를 묻는 것이 자연스럽다. 각각의 시간에 1달러를 걸었을 경우 시간 n에 버는 돈의 총량을 X_n이라 하자. 도박사가 도박 체계 H_n에 의해 돈을 걸 경우 시간 n에서의 상금은 다음과 같다.

$$(H \cdot X)_n = \sum_{m=1}^{n} H_m(X_m - X_{m-1})$$

왜냐하면 시간 m에 도박사가 3달러를 걸었다면 도박사의 상금의 변화량은 1달러를 건 사람에 비해 3배가 될 것이기 때문이다. 아니면 X_m을 주가, H_m을 시간 $m - 1$부터 m까지의 기간 동안 보유하고 있는 주식의 수량이라 간주할 수 있다.

이제 $\xi_m = X_m - X_{m-1}$이 $P(\xi_m = 1) = p, P(\xi_m = -1) = 1 - p$를 가진다 하자. '마팅게일'이라 불리는 유명한 도박 체계는 $H_1 = 1$ 및 $n \ge 2$의 경우 다음 공식에 의해 정의된다.

$$H_n = \begin{cases} 2H_{n-1} & (\xi_{n-1} = -1) \\ 1 & (\xi_{n-1} = 1) \end{cases}$$

즉 게임에 지면 판돈을 2배로 올리며, 따라서 k회 진 후 이긴다면 총 상금이 1이 될 것이다. 이를 보이기 위해 다음의 구체적 상황을 고려하자.

H_n	1	2	4	8	16
ξ_n	-1	-1	-1	-1	1
$(H \cdot X)_n$	-1	-3	-7	-15	1

$P(\xi_m = 1) > 0$인 한 이러한 계는 '확실한 이득'을 제공하는 것처럼 보인다. 하지만 다음 예시는 불리한 게임을 이길 수 있는 방법이 없음을 보여준다.

Theorem 4.2.8. X_n $(n \geq 0)$이 우마팅게일이라 하자. 만약 $H_n \geq 0$이 예측 가능하며 각각의 H_n이 유계이면 $(H \cdot X)_n$이 우마팅게일이다.

Proof. 조건부기댓값이 선형이라는 사실을 이용하면 $(H \cdot X)_n \in \mathcal{F}_n, H_n \in \mathcal{F}_{n-1}$이므로 (4.1.14)에 의해 다음이 성립한다.

$$\begin{aligned} E((H \cdot X)_{n+1}|\mathcal{F}_n) &= (H \cdot X)_n + E(H_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n) \\ &= (H \cdot X)_n + H_{n+1}E((X_{n+1} - X_n)|\mathcal{F}_n) \leq (H \cdot X)_n \end{aligned}$$

여기에서 $E((X_{n+1} - X_n)|\mathcal{F}_n) \leq 0$이며 $H_{n+1} \geq 0$이라는 사실을 사용했다. □

Remark. 동일한 결과가 열마팅게일 및 마팅게일에 대해서도 자명하게 성립한다. (마지막 경우에는 $H_n \geq 0$이라는 제한이 필요하지 않다.)

우리는 이제 매우 특수한 도박 체계를 고려할 것이다: $n \leq N$인 각각의 시간에 1달러를 걸고 그 후 게임을 정지한다. 확률변수 N이 **종료시간(stopping time)**이라는 것의 정의는 모든 $n < \infty$에 대하여 $\{N = n\} \in \mathcal{F}_n$인 것이다. i.e. 시간 n에서 정지할지의 결정은 그 시각에 알고 있는 정보들에 대하여 가측이어야 한다. 만약 $H_n = 1_{N \geq n}$이라 하면 $\{N \geq n\} = \{N \leq n-1\}^c \in \mathcal{F}_{n-1}$이며 따라서 H_n이 예측 가능하다. Theorem 4.2.8로부터 $(H \cdot X)_n = X_{N \wedge n} - X_0$이 우마팅게일임이 따라온다. 상수 열 $Y_n = X_0$이 우마팅게일이며 두 우마팅게일의 합이 우마팅게일이므로 다음이 성립한다:

Theorem 4.2.9. 만약 N이 종료시각이며 X_n이 우마팅게일이면 $X_{N \wedge n}$이 우마팅게일이다.

Theorem 4.2.8이 도박 체계를 통해 돈을 벌 수 없음을 함의하지만, 이를 통해 정리들을 증명할 수 있다. $X_n, n \geq 0$이 열마팅게일이라 하자. $a < b$이며 $N_0 = -1$이고 $k \geq 1$에 대하여 다음과 같다 하자.

$$\begin{aligned} N_{2k-1} &= \inf\{m > N_{2k-2} : X_m \leq a\} \\ N_{2k} &= \inf\{m > N_{2k-1} : X_m \geq b\} \end{aligned}$$

N_j들은 종료시각이며 $\{N_{2k-1} < m \leq N_{2k}\} = \{N_{2k-1} \leq m-1\} \cap \{N_{2k} \leq m-1\}^c \in \mathcal{F}_{m-1}$이며, 따라서

$$H_m = \begin{cases} 1 & (\exists k \ N_{2k-1} < m \leq N_{2k}) \\ 0 & (\nexists k \ N_{2k-1} < m \leq N_{2k}) \end{cases}$$

는 예측 가능 열을 정의한다. $X(N_{2k-1}) \leq a, X(N_{2k}) \geq b$이므로 N_{2k-1}과 N_{2k} 사이의 시간에서 X_m은 a 아래에서 b 위로 넘어간다. H_m은 이러한 '상승교차'의 이점을 취하려 시도하는 도박 체계이다. 주식시장의 용어로 말하면, $X_m \leq a$일 때 매수하여 $X_m \geq b$일 때 매도하면 상승교차가 완료된 시점에서 $b-a$ 이상의 이윤을 얻는다. 마지막으로 $U_n = \sup\{k : N_{2k} \leq n\}$은 시간 n에 완료된 상승교차의 개수이다.

Theorem 4.2.10 (Upcrossing inequality (상승교차 부등식))**.** 만약 X_m $(m \geq 0)$이 열마팅게일이면 다음이 성립한다.

$$(b-a)EU_n \leq E(X_n - a)^+ - E(X_0 - a)^+$$

Proof. $Y_m = a + (X_m - a)^+$라 하자. Theorem 4.2.7에 의해 Y_m은 열마팅게일이다. 명백히 이는 X_m이 $[a, b]$를 상승교차하는 횟수와 동일한 횟수만큼 상승교차한다. 각각의 상승교차에서 $b-a$ 이상의 이윤을 얻으며 마지막의 불완전한 상승교차는 (존재한다면) $(H \cdot Y)_n$에 음 아닌 수만큼 기여하므로 $(b-a)U_n \leq (H \cdot Y)_n$이 성립한다. (이 부등식을 얻기 위해 X_m을 Y_m으로 대체했다.)

$K_m = 1 - H_m$이라 하자. 명백히 $Y_n - Y_0 = (H \cdot Y)_n + (K \cdot Y)_n$이며 Theorem 4.2.8로부터 $E(K \cdot Y)_n \geq E(K \cdot Y)_0 = 0$이며 따라서 $E(H \cdot Y)_n \leq E(Y_n - Y_0)$임이 따라온다. 이는 요구된 부등식을 증명한다. □

우리는 위 결과를 (약간의 혼동의 여지가 있는) 고전적 형태로 증명했다. 핵심적인 사실은 $E(K \cdot Y)_n \geq 0$이라는 것이다. i.e. 도박사가 아무리 노력하더라도 열마팅게일에 돈을 거는 것으로는 돈을 잃을 수 없다. 상승교차 부등식으로부터 다음을 간단히 얻는다:

Theorem 4.2.11 (Martingale convergence theorem (마팅게일 수렴 정리))**.** 만약 X_n이 $\sup EX_n^+ < \infty$를 만족시키는 열마팅게일이면 $n \to \infty$에서 X_n은 $E|X| < \infty$를 만족시키는 극한 X로 a.s. 수렴한다.

Proof. $(X-a)^+ \le X^+ + |a|$이므로 THeorem 4.2.10은 다음을 함의한다.

$$EU_n \le (|a| + EX_n^+)/(b-a)$$

$n \uparrow \infty$에서 $U_m \uparrow U$ (전체 수열에서 $[a,b]$의 상승교차 횟수)로 수렴한다. 그러므로 만약 $\sup EX_n^+ < \infty$이면 $EU < \infty$이며 따라서 $U < \infty$ a.s.이다. 이 결과가 모든 유리수 a, b에 대하여 성립하므로 다음이 성립한다.

$$\bigcup_{a,b \in \mathbb{Q}} \{\liminf X_n < a < b < \limsup X_n\} \quad \text{가 확률 0을 가진다.}$$

따라서 $\limsup X_n = \liminf X_n$ a.s.이다. i.e. $\lim X_n$이 a.s. 존재한다. Fatou 보조정리는 $EX^+ \le \liminf EX_n^+ \le \infty$임을 보장하며 따라서 $X < \infty$ a.s.이다. $X > -\infty$임을 보이기 위해서는 (X_n이 열마팅게일이므로) 다음이 성립함을 관찰하라.

$$EX_n^- = EX_n^+ - EX_n \le EX_n^+ - EX_0$$

그러므로 Fatou 보조정리를 다시 적용하면 다음을 얻는다.

$$EX^- \le \liminf_{n \to \infty} EX_n^- \le \sup_n EX_n^+ - EX_0 < \infty$$

이는 증명을 완료한다. □

Remark. Theorem 4.7.1의 증명을 준비하기 위해 독자는 모든 $a, b \in \mathbb{Q}$에 대하여 X_n에 의한 (a, b)의 상승교차수가 유한하면 X_n의 극한이 존재함을 기억해 두어야 한다.

Theorem 4.2.11의 중요한 특수한 경우는 다음과 같다:

Theorem 4.2.12. 만약 $X_n \ge 0$이 우마팅게일이면 $n \to \infty$에서 $X_n \to X$ a.s.이며 $EX \le EX_0$이다.

Proof. $Y_n = -X_n \le 0$은 $EY_n^+ = 0$을 만족시키는 열마팅게일이다. $EX_0 \ge EX_n$이므로 부등식은 Fatou 보조정리에서 따라온다. □

다음 절에서 우리는 마지막 두 결과들의 여러 응용을 제시할 것이다. 두 가지 '반례'를 제시하는 것으로 이 절을 마치겠다.

Example 4.2.13. Theorem 4.2.12(또는 4.2.11)의 결과는 L^1에서의 수렴성을 보장하지 않는다. S_n이 $S_0 = 1$에서 시작하는 대칭 단순 난보라 하자. i.e. $S_n = S_{n-1} + \xi_n$이며 $\xi_1, \xi_2, \ldots$는 $P(\xi_i = 1) = P(\xi_i = -1) = 1/2$을 가지는 i.i.d.이다. $N = \inf\{n : S_n = 0\}$이며 $X_n = S_{N \wedge n}$이라 하자. Theorem 4.2.9는 X_n이 음 아닌 마팅게일임을 함의한다. Theorem 4.2.12는 X_n이 극한 $X_\infty < \infty$로 수렴함을 함의하며, $k > 0$으로의 수렴이 불가능하므로 이러한 극한은 항등적으로 0이어야 한다. (만약 $X_n = k > 0$이면 $X_{n+1} = k \pm 1$이다.) 모든 n에 대하여 $EX_n = EX_0 = 1$이며 $X_\infty = 0$이므로 L^1에서는 수렴이 발생할 수 없다.

Example 4.2.13은 이 장의 남은 부분을 읽으면서 명심하고 있어야 할 중요한 반례이다. 다음 것은 그만큼 중요하지는 않다.

Example 4.2.14. $X_k \to 0$으로 확률에서 수렴하지만 a.s. 수렴하지는 않는 마팅게일의 예시를 제시하겠다. $X_0 = 0$이라 하자. $X_{k-1} = 0$일 경우 $X_k = 1$ 또는 -1일 확률이 각각 $1/2k$이며 0일 확률이 $1 - 1/k$라 하자. $X_{k-1} \ne 0$일 경우 $1/k$의 확률로 $X_k = kx_{k-1}$이며 $1 - 1/k$의 확률로 0이라 하자. 구축에 의해 $P(X_k = 0) = 1 - 1/k$이며 따라서 $X_k \to 0$으로 확률에서 수렴한다. 반면에 제2 Borel-Cantelli 보조정리는 $P(k \ge K$에 대하여 $X_k = 0) = 0$임을 함의하며 $(-1, 1) - \{0\}$에 속한 값을 가질 수 없으므로 X_k는 0으로 a.s. 수렴하지 않는다.

Exercises (연습문제)

4.3 Examples (예시)

이 절에서 우리는 마팅게일 수렴 정리를 적용하여 제2 Borel-Cantelli 보조정리를 일반화하고 Polya 항아리 구조, Radon-Nikodym 도함수, 분기 과정을 다룰 것이다. 4가지 주제들은 서로 독립적이며 위 순서대로 다룰 것이다.

4.3.1 Bounded Increments (유계 증가)

우리의 첫째 결과는 유계 증가 마팅게일이 수렴하거나 또는 $+\infty$와 $-\infty$ 사이에서 진동함을 보여준다.

Theorem 4.3.1. $X_1, X_2, \ldots$가 마팅게일이며 $|X_{n+1} - X_n| \le M < \infty$라 하자. 다음이 성립한다 하자.

$$C = \{\lim X_n \text{이 존재하며 유한하다.}\}$$
$$D = \{\limsup X_n = +\infty\text{이며 } \liminf X_n = -\infty\}$$

그 경우 $P(C \cup D) = 1$이다.

Proof. $X_n - X_0$이 마팅게일이므로 일반성을 잃지 않고 $X_0 = 0$이라 가정할 수 있다. $0 < K < \infty$라 하고 $N \inf\{n : X_n \le -K\}$라 하자. $X_{n\wedge N}$은 $X_{n\wedge N} \ge -K-M$ a.s.를 만족시키는 마팅게일이므로 $X_{n\wedge N}+K+M$에 Theorem 4.2.12를 적용하면 $\{N = \infty\}$에서 $\lim X_n$이 존재함을 보여준다. $K \to \infty$라 하면 $\{\liminf X_n > -\infty\}$에서 극한이 존재함을 알 수 있다. 마지막 결론을 $-X_n$에 적용하면 $\{\limsup X_n < \infty\}$에서 $\lim X_n$이 존재함을 알 수 있다. 이는 증명을 완료한다. □

이 결과를 응용하기 위해서는 다음이 필요하다:

Theorem 4.3.2 (Doob's decomposition (Doob 분해))**.** 임의의 열마팅게일 X_n $(n \ge 0)$은 마팅게일 M_n과 $A_0 = 0$을 만족시키는 예측 가능 증가 열 A_n의 합 $X_n = M_n + A_n$으로 유일하게 표현 가능하다.

Proof. 우리는 $X_n = M_n + A_n, E(M_n|\mathcal{F}_{n-1}) = M_{n-1}, A_n \in \mathcal{F}_{n-1}$를 원하며, 따라서 다음이 필요하다.

$$\begin{aligned} E(X_n|\mathcal{F}_{n-1}) &= E(M_n|\mathcal{F}_{n-1}) + E(A_n|\mathcal{F}_{n-1}) \\ &= M_{n-1} + A_n = X_{n-1} - A_{n-1} + A_n \end{aligned}$$

그러므로 다음이 필요하다.

$$A_n - A_{n-1} = E(X_n|\mathcal{F}_{n-1}) - X_{n-1} \tag{4.3.1}$$

$A_0 = 0$이므로 다음이 성립해야 한다.

$$A_n = \sum_{m=1}^{n} E(X_n - X_{n-1}|\mathcal{F}_{n-1}) \tag{4.3.2}$$

이러한 A_n이 요구된 조건을 만족시킴을 확인하기 위해 X_n이 열마팅게일이며 $A_n \in \mathcal{F}_{n-1}$이므로 $A_n - A_{n-1} \ge 0$임을 관찰하라. $M_n = X_n - A_n$이 마팅게일임을 증명하기 위해 $A_n \in \mathcal{F}_{n-1}$임을 기억해 두고 (4.3.1)을 사용하라:

$$\begin{aligned} E(M_n|\mathcal{F}_{n-1}) &= E(X_n - A_n|\mathcal{F}_{n-1}) \\ &= E(X_n|\mathcal{F}_{n-1}) - A_n = X_{n-1} - A_{n-1} = M_{n-1} \end{aligned}$$

이는 증명을 완료한다. □

이러한 결과의 활용을 보여주기 위해 다음의 중요한 예시를 다루겠다.

Example 4.3.3. $B_n \in \mathcal{F}_n$이라 하자. (4.3.2)를 사용하면,

$$M_n = \sum_{m=1}^{n} 1_{B_m} = E(1_{B_m}|\mathcal{F}_{m-1})$$

Theorem 4.3.4 (Second Borel-Cantelli lemma, II (제2 Borel-Cantelli 보조정리, II))**.** $\mathcal{F}_n$ $(n \ge 0)$이 여과이며 $\mathcal{F}_0 = \{\emptyset, \Omega\}$이고 A_n $(n \ge 1)$이 $B_n \in \mathcal{F}_n$을 만족시키는 사건들의 열이라 하자. 그 경우,

$$\{B_n \text{ i.o.}\} = \left\{\sum_{n=1}^{\infty} P(B_n|\mathcal{F}_{n-1}) = \infty\right\}$$

Proof. 만약 $X_0 = 0, X_n = \sum_{m\le n} 1_{B_m}$이라 하면 X_n은 열마팅게일이다. (4.3.2)는 $B_n = \sum_{m=1}^{n} E(1_{A_m}|\mathcal{F}_{m-1})$임을 함의한다. 따라서 만약 $M_0 = 0$이며 $n \ge 1$에 대하여 다음이 성립하면,

$$M_n = \sum_{m=1}^{n} 1_{A_m} - P(A_m|\mathcal{F}_{m-1})$$

M_n이 마팅게일이며 $|M_n - M_{n-1}| \le 1$이 성립한다. Theorem 4.3.1의 표기법을 사용하면 다음을 얻는다:

$$\begin{aligned} C \text{ 상에서} \quad & \sum_{n=1}^{\infty} 1_{B_n} = \infty \quad \text{iff} \quad \sum_{n=1}^{\infty} P(B_n|\mathcal{F}_{n-1}) = \infty \\ D \text{ 상에서} \quad & \sum_{n=1}^{\infty} 1_{B_n} = \infty \quad \text{and} \quad \sum_{n=1}^{\infty} P(B_n|\mathcal{F}_{n-1}) = \infty \end{aligned}$$

$P(C \cup D) = 1$이므로 결과가 따라온다. □

4.3.2 Polya's Urn Scheme (Polya 항아리 구조)

항아리에 r개 적색 공, g개 녹색 공이 담겨있다 하자. 공을 하나씩 뽑을 때마다 이를 다시 넣고 뽑은 색 공을 c개 추가한다 하자. X_n이 n회 시행 이후 녹색 공의 비율이라 하자. X_n이 마팅게일임을 확인하기 위해 시간 n에 i개 적색 공과 j개 녹색 공이 존재하면 다음이 성립함을 기억해 두라.

$$X_{n+1} = \begin{cases} (j+c)/(i+j+c) & (\text{확률 } j/(i+j)) \\ j/(i+j+c) & (\text{확률 } i/(i+j)) \end{cases}$$

따라서 다음이 성립한다.

$$\frac{j+c}{i+j+c} \cdot \frac{j}{i+j} + \frac{i}{i+j+c} \cdot \frac{i}{i+j} = \frac{(j+c+i)j}{(i+j+c)(i+j)} = \frac{j}{i+j}$$

$X_n \ge 0$이므로 Theorem 4.2.12는 $X_n \to X_\infty$ a.s.임을 함의한다. 극한의 확률분포를 계산하기 위해 다음을 관찰하자: (a) 앞의 m번의 시행에서 녹색 공을 뽑고 그 후 $l = n - m$번의 시행에서 적색 공을 뽑을 확률은 다음과 같다.

$$\frac{g}{g+r} \cdot \frac{g+c}{g+r+c} \cdots \frac{g+(m-1)c}{g+r+(m-1)c} \cdot \frac{r}{g+r+mc} \cdots \frac{r+(l-1)c}{g+r+(n-1)c}$$

(b) n회의 시행에서 녹색 공 m개, 적색 공 l개를 뽑는 임의의 가능한 상황들은 (분모가 그대로이며 분자들이 치환되므로) 모두 이와 동일한 확률을 가진다. $c = 1, g = 1, r = 1$인 특수한 경우를 고려하자. G_n이 n번째 시행이 완료되어 새로운 공들이 추가된 이후의 녹색 공 개수라 하자. (a)와 (b)로부터 다음이 따라온다.

$$P(G_n = m+1) = \binom{n}{m} \frac{n!(n-m)!}{(n+1)!} = \frac{1}{n+1}$$

그러므로 X_∞는 $(0,1)$에서 균등 확률분포를 가진다.

만약 $c = 1, g = 2, r = 1$이라 하면 $n \to \infty, m/n \to x$에서 다음이 성립한다.

$$P(G_n = m+2) = \frac{n!}{m!(n-m)!} \frac{(m+1)!(n-m)!}{(n+2)!/2} \to 2x$$

일반적으로 X_∞의 확률분포는 다음의 확률밀도함수를 가진다.

$$\frac{\Gamma((g+r)/c)}{\Gamma(g/c)\Gamma(r/c)} x^{(g/c)-1}(1-x)^{(r/c)-1}$$

이는 매개변수 g/c와 r/c를 가지는 **베타 확률분포(beta distribution)**이다. Example 4.5.6에서 우리는 뽑힌 색의 공 c개에 더해 반대 색의 공 1개를 더 넣는다면 극한의 거동이 대폭 변화함을 보게 될 것이다.

4.3.3 Radon-Nikodym Derivatives (Radon-Nikodym 도함수)

μ가 유한 측도이며 ν가 $(\Omega, \mathcal{F})$에서의 확률측도라 하자. $\mathcal{F}_n \uparrow \mathcal{F}$가 σ-체들이라 하자. (i.e. $\sigma(\bigcup \mathcal{F}_n) = \mathcal{F}$.) μ_n과 ν_n이 각각 μ, ν의 $\mathcal{F}_n$으로의 제한이라 하자.

Theorem 4.3.5. 모든 n에 대하여 $\mu_n \ll \nu_n$이라 하자. $X_n = d\mu_n/d\nu_n$이며 $X = \limsup X_n$이라 하자. 그 경우,

$$\mu(A) = \int_A X\ d\nu + \mu(A \cap \{X = \infty\})$$

Remark. $\mu_r(A) = \int_A X\ d\nu$는 ν에 대하여 절대연속한 측도이다. Theorem 4.2.12가 $\nu(X = \infty)$임을 함의하므로 $\mu_s(A) \equiv \mu(A \cap \{X = \infty\})$는 ν에 대하여 특이이다. 그러므로 $\mu = \mu_r + \mu_s$는 μ의 Lebesgue 분해를 제공한다. (Theorem A.4.7을 참조하라.) 또한 $X_\infty = d\mu_r/d\nu, \nu$-a.s.이다. 이곳과 증명에서 우리는 a.s.가 어떤 측도에 대한 것인지를 계속 추적하겠다.

Proof. 독자가 예측할 수 있듯이 다음이 성립한다:

Lemma 4.3.6. $(\Omega, \mathcal{F}, \nu)$에서 정의된 X_n은 $\mathcal{F}_n$에 대한 마팅게일이다.

Proof. 정의에 의해 $X_n \in \mathcal{F}_n$임을 관찰하라. $A \in \mathcal{F}_n$이라 하자. $X_n \in \mathcal{F}_n$이며 ν_n이 ν의 $\mathcal{F}_n$으로의 제한이므로,

$$\int_A X_n\ d\nu = \int_A X_n\ d\nu_n$$

X_n의 정의와 Exercise A.4.7을 사용하면 다음이 성립한다.

$$\int_A X_n\ d\nu_n = \mu_n(A) = \mu(A)$$

($A \in \mathcal{F}_n$이며 μ_n이 μ의 $\mathcal{F}_n$으로의 제한이므로 마지막 등식이 성립한다.) 만약 $A \in \mathcal{F}_{m-1} \subset \mathcal{F}_m$인 경우 $n = m$과 $n = m - 1$에 대하여 마지막 결과를 적용하면 다음을 얻는다.

$$\int_A X_m\ d\nu = \mu(A) = \int_A X_{m-1}\ d\nu$$

따라서 $E(X_m|\mathcal{F}_{m-1}) = X_{m-1}$이다. □

X_n이 음 아닌 마팅게일이므로 Theorem 4.2.12는 $X_n \to X$ ν-a.s.임을 함의한다. 우리는 정리에서 등호가 성립함을 보이고자 한다. $\mu(A)$를 $\mu(\Omega)$로 나누면 일반성을 잃지 않고 μ가 확률측도라 가정할 수 있다. $\rho = (\mu + \nu)/2, \rho_n = (\mu_n + \nu_n)/2 = \rho$의 $\mathcal{F}_n$으로의 제한이라 하자. $Y_n = d\mu_n/d\rho_n, Z_n = d\nu_n/d\rho_n$이라 하자. $Y_n, Z_n \geq 0$이며 (Exercise 2.4.6에 의해) $Y_n + Z_n = 2$이므로 Y_n과 Z_n은 극한 Y, Z를 가지는 유계 마팅게일이다. 독자가 유추할 수 있는 것처럼 다음이 성립한다.

$$Y = d\mu/d\rho \quad , \quad Z = d\nu/d\rho \tag{*}$$

첫째 등식을 보이면 충분하다. Lemma 4.3.6의 증명에 의해 만약 $A \in \mathcal{F}_m \subset \mathcal{F}_n$이면 유계 수렴 정리에 의해 다음이 성립한다.

$$\mu(A) = \int_A Y_n\ d\rho \to \int_A Y\ d\rho$$

마지막 계산은 다음을 보여준다.

$$\mu(A) = \int_A Y\ d\rho \qquad \left(\forall A \in \mathcal{G} = \bigcup_m \mathcal{F}_m\right)$$

$\mathcal{G}$가 π-계이므로 π-λ 정리에 의해 모든 $A \in \mathcal{F} = \sigma(\mathcal{G})$에 대하여 등호가 성립하며 (*)가 증명되었다.

Exercise A.4.8과 A.4.9로부터 $X_n = Y_n/Z_n$임이 따라온다. 이로부터 독자는 아마도 $X = Y/Z$라는 결론을 추론할 수 있을 것이다. 이러한 결론에 도달하기 위해 $Y + Z = 2$ ρ-a.s.이며 따라서 $\rho(Y = 0, Z = 0) = 0$임을 기억해 두라. 0/0을 제외하면 $X = Y/Z$ ρ-a.s.임을 얻는다. ($X \equiv \limsup X_n$임을 상기하라.) $W = (1/Z) \cdot 1_{(Z>0)}$이라 하자. (*)를 사용하면 $1 = ZW + 1_{(Z=0)}$이며 다음이 성립한다.

$$\mu(A) = \int_A Y\ d\rho = \int_A YWZ\ d\rho + \int_A 1_{(Z=0)} Y\ d\rho \tag{a}$$

이제 (*)는 $d\nu = Z\ d\rho$임을 함의하며 정의로부터 다음이 따라온다.

$$YW = X1_{(Z>0)} = X \quad \nu\text{-a.s.}$$

($\nu(\{Z=0\}) = 0$이므로 둘째 등식이 성립한다.) 이들을 조합하면 다음을 얻는다.

$$\int_A YWZ\ d\rho = \int_A X\ d\nu \tag{b}$$

다른 항을 다루기 위해 (*)가 $d\mu = Y\ d\rho$임을 함의함을 기억해 두라. 정의로부터 $\{X = \infty\} = \{Z = 0\}$ μ-a.s.임이 따라온다. 따라서,

$$\int_A 1_{(Z=0)} Y\ d\rho = \int_A 1_{(X=\infty)}\ d\mu \tag{4.3.3}$$

(a), (b), (c)를 조합하면 요구된 결과를 얻는다. □

Example 4.3.7. $\mathcal{F}_n = \sigma(I_{k,n} : 0 \le k < K_n)$이라 가정하자. 여기에서 각각의 n에 대하여 $I_{k,n}$은 Ω의 분할이며 $(n+1)$번째 분할은 n번째 분할의 세분할이다. 이 경우 $\mu_n \ll \nu_n$이라는 조건은 $\nu(I_{k,n}) = 0$이면 $\mu(I_{k,n}) = 0$임을 함의한다. $I_{k,n}$에서 $X_n = \mu(I_{k,n})/\nu(I_{k,m})$으로 정의된 마팅게일 X_n은 Radon-Nikodym 도함수에 대한 접근이다. 구체적인 예시로, $\Omega = [0,1), I_{k,n} = [k2^{-n}, (k+1)2^{-n})\ (0 \le k < 2^n), \nu =$Lebesgue 측도인 경우를 고려하라.

Kakutani dichotomy for infinite product measures (무한 곱측도에 대한 Kakutani 이분법). μ와 ν가 점렬공간 $(\mathbb{R}^{\mathbb{N}}, \mathbb{R}^{\mathbb{N}})$에서의 측도이며 이들 하에서 좌표 $\xi_n(\omega) = \omega_n$들이 독립이라 하자. $F_n(x) = \mu(\xi_n \le x), G_n(x) = \nu(\xi_n \le x)$라 하자. $F_n \ll G_n$이며 $q_n = dF_n/dG_n$이라 하자. 문제를 회피하기 위해서는 $q_n > 0, G_n$-a.s.라 가정해야 한다.

$\mathcal{F}_n = \sigma(\xi_m : m \le n)$이며 μ_n, ν_n이 각각 μ, ν의 $\mathcal{F}_n$으로의 제한을 나타낸다 하자. 또한 다음과 같다 하자.

$$X_n = \frac{d\mu_n}{d\nu_n} = \prod_{m=1}^{n} q_m$$

Theorem 4.3.5는 $X_n \to X$ ν-a.s. 임을 함의한다. 가정에 의해 $q_n > 0, G_n$-a.s.이므로 $\sum_{m=1}^{\infty} \log(q_m) > -\infty$는 꼬리사건이며 따라서 Kolmogorov 0-1 법칙은 다음을 함의한다.

$$\nu(X = 0) \in \{0, 1\} \tag{4.3.4}$$

Theorem 4.3.5로부터 $\mu \ll \nu$ 또는 $\mu \perp \nu$임이 따라온다. 다음 결과는 두 가지 서로 다른 상황이 언제 발생하는지에 대한 구체적인 판정법을 제공한다.

Theorem 4.3.8. $\prod_{m=1}^{\infty} \int \sqrt{q_m}\ dG_m > 0$인지 $= 0$인지에 따라 $\mu \ll \nu$ 또는 $\mu \perp \nu$가 결정된다.

Proof. Jensen 부등식과 Exercise A.4.7은 다음을 함의한다.

$$\left(\int \sqrt{q_m}\ dG_m \right)^2 \le \int q_m\ dG_m = \int dF_m = 1$$

그러므로 적분들의 무한합이 잘 정의되며 1 이하이다. 위에서와 같이 다음과 같다 하자.

$$X_n = \prod_{m \le n} q_m(\omega_m)$$

$X_n \to X$ ν-a.s.임을 상기하라. 만약 무한곱이 0이면 다음이 성립한다.

$$\int X_n^{1/2}\ d\nu = \prod_{m=1}^{n} \int \sqrt{q_m}\ dG_m \to 0$$

Fatou 보조정리를 적용하면 다음을 얻는다.

$$\int X^{1/2}\ d\nu \le \liminf_{n \to \infty} \int X_n^{1/2}\ d\nu = 0$$

그러므로 $X = 0$ ν-a.s.이며 따라서 Theorem 4.3.5는 $\mu \perp \nu$임을 함의한다. 다른 방향을 증명하기 위해 $Y_n = X_n^{1/2}$라 하자. 이제 $\int q_m \, dG_m = 1$이므로 (E가 ν에 대한 기댓값을 나타낸다 하면) $EY_m^2 = EX_m = 1$이다. 따라서 다음이 성립한다.

$$E(Y_{n+k} - Y_n)^2 = E(X_{n+k} + X_n - 2X_n^{1/2}X_{n+k}^{1/2}) = 2\left(1 - \prod_{m=n+1}^{n+k} \int \sqrt{q_m} \, dG_m\right)$$

이제 $|a-b| = |a^{1/2} - b^{1/2}| \cdot (a^{1/2} + b^{1/2})$이며 따라서 Cauchy-Schwarz 및 $(a+b)^2 \le 2a^2 + 2b^2$라는 사실을 사용하면 다음을 얻는다.

$$\begin{aligned} E|X_{n+k} - X_n| &= E(|Y_{n+k} - Y_n|(Y_{n_k} - Y_n)) \\ &\le \left(E(Y_{n+k} - Y_n)^2 E(Y_{n+k} + Y_n)^2\right)^{1/2} \\ &\le \left(4E(Y_{n+k} - Y_n)^2\right)^{1/2} \end{aligned}$$

마지막 두 방정식에서 무한곱이 0 초과이면 $L^1(\nu)$에서 X_n이 X로 수렴하며 따라서 $\nu(X = 0) < 1$이고 (4.3.3)에 의해 확률이 0임이 따라온다. 요구된 결과는 Theorem 4.3.5에서 따라온다. □

4.3.4 Branching Processes (분기 과정)

ξ_i^n $(i, n \ge 1)$이 i.i.d. 음 아닌 정수 값 확률변수라 하자. 열 Z_n $(n \ge 0)$을 $Z_0 = 1$과 다음 식에 의해 정의하자.

$$Z_{n+1} = \begin{cases} \xi_1^{n+1} + \cdots + \xi_{Z_n}^{n+1} & (Z_n > 0) \\ 0 & (Z_n = 0) \end{cases} \tag{4.3.5}$$

Z_n은 **Galton-Watson 과정(Galton-Watson process)**이라 불린다. 정의의 기반이 되는 발상은 Z_n이 n번째 세대의 개체수이며 n번째 세대의 각 구성원은 독립 동분포 명수의 아이들을 낳는다는 것이다. $p_k = P(\xi_i^n = k)$는 **자손분포(offspring distribution)**라 불린다.

Lemma 4.3.9. $\mathcal{F}_n = \sigma(\xi_i^m : i \ge 1, 1 \le m \le n)$이며 $\mu = E\xi_i^n \in (0, \infty)$라 하자. 그 경우 Z_n/μ^n은 $\mathcal{F}_n$에 대한 마팅게일이다.

Proof. 명백히 $Z_n \in \mathcal{F}_n$이다. Theorem 4.1.2를 사용하여 $\{Z_n = k\}$에서 다음이 성립한다 결론지을 수 있다.

$$E(Z_{n+1}|\mathcal{F}_n) = E(\xi_1^{n+1} + \cdots + \xi_k^{n+1}|\mathcal{F}_n) = k\mu = \mu Z_n$$

둘째 등식에서 ξ_k^{n+1}이 $\mathcal{F}_n$에 독립이라는 사실을 사용했다. □

Z_n/μ^n이 음 아닌 마팅게일이므로 Theorem 4.2.12는 Z_n/μ^n이 어떠한 극한으로 a.s. 수렴함을 함의한다. 극한이 자명한 경우를 식별하는 것으로 시작하겠다.

Theorem 4.3.10. 만약 $\mu < 1$이면 충분히 큰 모든 n에 대하여 $Z_n = 0$이며 따라서 $Z_n/\mu^n \to 0$이다.

Proof. $E(Z_n/\mu^n) = E(Z_0) = 1$이며 따라서 $E(Z_n) = \mu^n$이다. 이제 $\{Z_n > 0\}$에서 $Z_n \ge 1$이며 따라서 ($\mu < 1$이므로) 다음이 지수적으로 수렴한다.

$$P(Z_n > 0) \le E(Z_n; Z_n > 0) = E(Z_n) = \mu^n \to 0$$

□

마지막 답은 직관적이다. 만약 평균적인 개인이 1명보다 적은 아이를 낳는다면 종이 멸종할 것이다. 다음 결과는 $\mu = 1$일 경우 모든 개인이 정확히 한 명의 아이를 낳는 자명한 경우를 제외하면 위와 동일한 결과가 성립함을 보여줄 것이다.

Theorem 4.3.11. 만약 $\mu = 1$이며 $P(\xi_i^m = 1) < 1$이므로 충분히 큰 모든 n에 대하여 $Z_n = 0$이다.

Proof. $\mu = 1$일 경우 Z_n 자신이 음 아닌 마팅게일이다. Z_n이 정수 값을 가지며 Theorem 4.2.12에 의해 유한한 극한 Z_∞로 a.s. 수렴하므로 충분히 큰 n에 대하여 $Z_n = Z_\infty$여야 한다. 만약 $P(\xi_i^m = 1) < 1$이며 $k > 0$이면 임의의 N에 대하여 $P($모든 $n \ge N$에 대하여 $Z_n = k) = 0$이며 따라서 $Z_\infty \equiv 0$이어야 한다. □

$\mu \le 1$일 경우 분기 과정이 점점 사라지므로 Z_n/μ^n의 극한은 0이다. 다음 단계는 $\mu > 1$일 경우 $P(\forall n\ Z_n > 0) > 0$임을 보이는 것이다. $p_k = P(\xi_i^m = k)$이며 $s \in [0,1]$에 대하여 $\varphi(s) = \sum_{k\ge 0} p_k s^k$라 하자. φ는 자손분포 p_k에 대한 **생성함수(generating function)**이다.

Theorem 4.3.12. $\mu > 1$이라 가정하자. 만약 $Z_0 = 1$이면 $P(\exists n\ Z_n = 0) = \rho$는 $[0,1)$에서 $\varphi(\rho) = \rho$의 유일한 해이다.

Proof. $\varphi(1) = 1$이다. 미분하면 $s < 1$에 대하여 다음을 얻는다. (이는 Theorem A.5.3에 의해 정당화된다.)

$$\varphi'(s) = \sum_{k=1}^{\infty} k p_k s^{k-1} \ge 0$$

따라서 φ는 증가함수이다. $s > 1$일 경우 $\varphi(s) = \infty$일 수 있으므로 주의하여 작업해야 한다.

$$\lim_{s\uparrow 1} \varphi'(s) = \sum_{k=1}^{\infty} k p_k = \mu$$

적분하면 $h \to 0$에서 다음을 얻는다.

$$\varphi(1) - \varphi(1-h) = \int_{1-h}^{1} \varphi'(s)\, ds \sim \mu h$$

그러므로 만약 h가 작다면 $\varphi(1-h) < 1-h$이다. $\varphi(0) \ge 0$이며 따라서 $[0,1)$에 $\varphi(x) = x$의 해가 존재해야 한다.

유일성을 증명하기 위해서는 $\mu > 1$이므로 어떠한 $k \ge 2$에 대하여 $p_k > 0$이며 따라서 $s < 1$에 대하여 다음이 성립함을 기억해 두라.

$$\varphi''(s) = \sum_{k=2}^{\infty} k(k-1) p_k s^{k-2} > 0$$

ρ가 $\varphi(\rho) = \rho$의 $[0,1)$에서의 최소해라 하자. $\varphi(1) = 1$이며 φ가 강 볼록이므로 $x \in (\rho, 1)$에 대하여 $\varphi(x) < x$이며 따라서 $[0,1)$에서 $\varphi(\rho) = \rho$의 해가 유일하게 존재한다.

다음 두 결과들을 조합하면 증명이 완료될 것이다.

(a) 만약 $\theta_m = P(Z_m = 0)$이면 $\theta_m = \sum_{k=0}^{\infty} p_k(\theta_{m-1})^k = \varphi(\theta_{m-1})$이다.

Proof of (a). 만약 $Z_1 = k$(확률 p_k의 사건)이면 $Z_m = 0$일 필요충분조건은 남은 $m-1$회에서 k개 족이 모두 사라지는 것이다. 이는 확률 θ_{m-1}^k를 가지는 독립 사건이다. 각각의 k에 대한 서로 소 가능성들을 모두 합하면 요구된 결과를 얻는다.

(b) $m \uparrow \infty$에서 $\theta_m \uparrow \rho$이다.

Proof of (b). 명백히 $\theta_m = P(Z_m = 0)$은 증가수열이다. 귀납법에 의해 $\theta_m \le \rho$이다: $\theta_0 = 0 \le \rho$이며 만약 이것이 $m-1$에 대하여 성립한다면 다음에 의해 m에 대해서도 성립한다.

$$\theta_m = \varphi(\theta_{m-1}) \le \varphi(\rho) = \rho$$

$\theta_m = \varphi(\theta_{m-1})$에 극한을 취하면 $\theta_\infty = \varphi(\theta_\infty)$를 얻는다. $\theta_\infty \le \rho$이므로 $\theta_\infty = \rho$임이 따라온다. □

마지막 결과는 $\mu > 1$일 경우 Z_n/μ^n의 극한이 0이 아닐 가능성이 있음을 보여준다 이러한 문제에 대한 최적의 결과는 Kesten과 Stigum에 의한 다음 결과이다:

Theorem 4.3.13. $W = \lim Z_n/\mu^n$이 항등적으로 0이지는 않음 iff $\sum p_k k \log k < \infty$인 것이다.

증명을 위해서는 Athreya and Ney (1972), pp. 24-29를 참조하라. 다음 절에서 우리는 $\sum k^2 p_k < \infty$가 비자명 극한에 대한 충분조건임을 보일 것이다.

4.4 Doob's Inequality, Convergence in $L^p, p > 1$ (Doob 부등식, $p > 1$인 L^p에서의 수렴)

Theorem 4.2.9의 결과를 증명하는 것으로 시작하겠다.

Theorem 4.4.1. 만약 X_n이 열마팅게일이며 N이 $P(N \le k) = 1$을 만족시키는 정지시간이면 다음이 성립한다.

$$EX_0 \le EX_N \le EX_k$$

Remark. S_n이 $S_0 = 1$을 만족시키는 단순난보이며 $N = \inf\{n : S_n = 0\}$이라 하자. (세부사항을 위해서는 Example 4.2.13을 참조하라.) $ES_0 = 1 > 0 = ES_N$이며 따라서 비유계 종료시간에 대하여 첫째 부등식이 성립할 필요는 없다. Section 5.7에서 우리는 비유계 N에 대하여 $EX_0 \le EX_N$임을 보장하는 조건을 제시할 것이다.

Proof. Theorem 4.2.9는 $X_{N\wedge n}$이 열마팅게일임을 함의하며, 따라서 다음이 성립한다.

$$EX_0 = EX_{N\wedge 0} \le EX_{N\wedge k} = EX_N$$

다른 방향 부등식을 증명하기 위해 $K_n = 1_{\{N<n\}} = 1_{\{N\le n-1\}}$이라 하자. K_n은 예측 가능하며 따라서 Theorem 4.2.8은 $(K \cdot X)_n = X_n - X_{N\wedge n}$이 열마팅게일임을 함의한다. 다음이 따라온다.

$$EX_k - EX_N = E(K \cdot X)_k \ge E(K \cdot X)_0 = 0$$

□

우리는 나중에 Theorem 4.4.1이 매우 유용함을 알게 될 것이다. 이것이 함의하는 바는 첫째로 다음이 있다:

Theorem 4.4.2 (Doob's inequality (Doob 부등식))**.** X_m이 열마팅게일이며 다음과 같다 하자.

$$\bar{X}_n = \max_{0\le m\le n} X_m^+$$

또한 $\lambda > 0$이며 $A = \{\bar{X}_n > \lambda\}$라 하자. 그 경우 다음이 성립한다.

$$\lambda P(A) \le EX_n 1_A \le EX_n^+$$

Proof. $N = \inf\{m : X_m \ge \lambda \text{ 또는 } m = n\}$이라 하자. A에서 $X_N \ge \lambda$이므로 다음이 성립한다.

$$\lambda P(A) \le EX_N 1_A \le EX_n 1_A$$

(여기에서 둘째 부등식은 Theorem 4.4.1에 의해 $EX_N \le EX_n$이며 A^c에서 $X_N = X_n$이므로 따라온다.) 둘째 부등식은 자명하므로 증명이 완료된다. □

Example 4.4.3 (Random walks (난보))**.** 만약 ξ_m들이 독립이고 $E\xi_m = 0, \sigma_m^2 = E\xi_m^2 < \infty$이며 $S_n = \xi_1 + \cdots + \xi_n$이라 하면 S_n은 마팅게일이다. 따라서 Theorem 4.2.6은 $X_n = S_n^2$가 열마팅게일임을 함의한다. $\lambda = x^2$라 하고 X_n에 Theorem 4.4.2를 적용하면 Kolmogorov 극대 부등식(Theorem 2.5.5)을 얻는다:

$$P\left(\max_{1\le m\le n} |S_m| \ge x\right) \le x^{-2}\operatorname{var}(S_n)$$

Theorem 4.4.2의 부등식을 적분하면 다음을 얻는다:

Theorem 4.4.4 (L^p 극대 부등식 (L^p maximum inequality))**.** 만약 X_n이 열마팅게일이면 $1 < p < \infty$에 대하여 다음이 성립한다.

$$E(\bar{X}_n^p) \le \left(\frac{p}{p-1}\right)^p E(X_n^+)^p$$

결과적으로 만약 Y_n이 마팅게일이며 $Y_n^* = \max_{0\le m\le n} |Y_m|$이면 다음이 성립한다.

$$E|Y_n^*|^p \le \left(\frac{p}{p-1}\right)^p E(|Y_n|^p)$$

Proof. 둘째 부등식은 첫째 부등식을 $X_n = |Y_n|$에 적용하면 따라온다. 첫째 부등식을 증명하기 위해 $\bar{X}_n$ 대신 $\bar{X}_n \wedge M$을 가지고 작업하겠다. $\{\bar{X}_n \wedge M \geq \lambda\}$는 항상 $\{\bar{X}_n \geq \lambda\}$ 또는 $\emptyset$이므로 이는 Theorem 4.4.2의 응용에 영향을 주지 않는다. Lemma 2.2.13, Theorem 4.4.2, Fubini 정리와 약간의 계산을 통해 다음을 얻는다.

$$\begin{aligned}
E((\bar{X}_n \wedge M)^p) &= \int_0^\infty p\lambda^{p-1} P(\bar{X}_n \wedge M \geq \lambda)\, d\lambda \\
&\leq \int_0^\infty p\lambda^{p-1} \left(\lambda^{-1} \int X_n^+ 1_{\bar{X}_n \wedge M \geq \lambda}\, dP\right) d\lambda \\
&= \int X_n^+ \int_0^{\bar{X}_n \wedge M} p\lambda^{p-2}\, d\lambda\, dP \\
&= \frac{p}{p-1} \int X_n^+ (\bar{X}_n \wedge M)^{p-1}\, dP
\end{aligned}$$

만약 $q = p/(p-1)$이 p의 공액 지수라 하고 Hölder 부등식(Theorem 1.6.3)을 적용하면 다음을 보일 수 있다.

$$\leq \left(\frac{p}{1-p}\right) (E|X_n^+|^p)^{1/p} (E|\bar{X}_n \wedge M|^p)^{1/q}$$

위 부등식의 양변을 $(E|\bar{X}_n \wedge M|^p)^{1/q}$ ($\wedge M$에 의해 이는 유한하다)로 나누고 양변에 p승멱을 취하면 다음을 얻는다.

$$E(|\bar{X}_n \wedge M|^p) \leq \left(\frac{p}{p-1}\right)^p E(X_n^+)^p$$

$M \to \infty$의 극한을 취하고 단조 수렴 정리를 사용하면 요구된 결과를 얻는다. □

Example 4.4.5 (There is no L^1 maximal inequality (L^1 극대 부등식의 부재성))**.** 반례는 다시 Example 4.2.13에 의해 제공된다. S_n이 $S_0 = 1$에서 시작하는 단순난보이며 $N = \inf\{n : S_n = 0\}, X_n = S_{N\wedge n}$이라 하자. Theorem 4.4.1은 모든 n에 대하여 $EX_n = ES_{N\wedge n} = ES_0 = 1$임을 함의한다. Theorem 4.8.7의 단순난보에 대한 도달확률을 사용하면 다음을 얻는다.

$$P\left(\max_m X_m \geq M\right) = \frac{1}{M} \tag{4.4.1}$$

따라서 $E(\max_m X_m) = \sum_{M=1}^\infty P(\max_m X_m \geq M) = \sum_{M=1}^\infty 1/M = \infty$이다. 단조 수렴 정리는 $n \uparrow \infty$에서 $E \max_{m\leq n} X_m \uparrow \infty$임을 함의한다.

Theorem 4.4.4로부터 다음을 얻는다.

Theorem 4.4.6 (L^p convergence theorem (L^p 수렴 정리))**.** 만약 X_n이 마팅게일이며 $p > 1$이고 $\sup E|X_n|^p < \infty$이면 $X_n \to X$로 a.s. 수렴하며 L^p 수렴한다.

Proof. $(EX_n^+)^p \leq (E|X_n|)^p \leq E|X_n|^p$이며 따라서 마팅게일 수렴 정리(Theorem 4.2.11)에서 $X_n \to X$ a.s.임이 따라온다. Theorem 4.4.4의 둘째 결론은 다음을 함의한다.

$$E\left(\sup_{0\leq m\leq n} |X_m|\right)^p \leq \left(\frac{p}{p-1}\right)^p E|X_n|^p$$

$n \to \infty$의 극한을 취하고 단조 수렴 정리를 사용하면 $\sup|X_n| \in L^p$임을 얻는다. $|X_n - X|^p \leq (2\sup|X_n|)^p$이므로 지배 수렴 정리에서 $E|X_n - X|^p \to 0$임이 따라온다. □

이 절의 결과들의 중요한 특수한 경우는 $p = 2$인 경우에 발생한다. 이 경우를 다루기 위해서는 다음 두 가지 결과들이 유용하다.

Theorem 4.4.7 (Orthogonality of martingale increments (마팅게일 증가량의 직교성))**.** X_n이 마팅게일이며 모든 n에 대하여 $EX_n^2 < \infty$를 만족시킨다 하자. 만약 $m \leq n$이며 $Y \in \mathcal{F}_n$이 $EY^2 < \infty$를 만족시키면 다음이 성립하며,

$$E((X_n - X_m)Y) = 0$$

따라서 $l < m < n$에 대하여 다음이 성립한다.

$$E((X_n - X_m)(X_m - X_l)) = 0$$

Proof. Cauchy-Schwarz 부등식은 $E|(X_n - X_m)Y| < \infty$임을 함의한다. (4.1.5), Theorem 4.1.14, 마팅게일의 정의를 사용하면 다음을 얻는다.

$$E((X_n - X_m)Y) = E[E((X_n - X_m)Y|\mathcal{F}_m)] = E[YE((X_n - X_m)|\mathcal{F}_m)] = 0$$

□

Theorem 4.4.8 (Conditional variance formula (조건부분산 공식))**.** 만약 X_n이 마팅게일이며 모든 n에 대하여 $EX_n^2 < \infty$를 만족시키면 다음이 성립한다.

$$E((X_n - X_m)^2|\mathcal{F}_m) = E(X_n^2|\mathcal{F}_m) - X_m^2$$

Remark. 이는 $E(X-EX)^2 = EX^2 - (EX)^2$의 조건부기댓값에 대한 유사품이며 완전히 동일한 방식으로 증명된다.

Proof. 조건부기댓값의 선형성과 Theorem 4.1.14를 사용하면 다음을 얻는다.

$$\begin{aligned}
E(X_n^2 - 2X_nX_m + X_m^2|\mathcal{F}_m) &= E(X_n^2|\mathcal{F}_m) - 2X_mE(X_n|\mathcal{F}_m) + X_m^2 \\
&= E(X_n^2|\mathcal{F}_m) - 2X_m^2 + X_m^2
\end{aligned}$$

이는 요구된 결과를 제공한다. □

Example 4.4.9 (Branching processes (분기 과정))**.** 앞 절의 마지막 부분에서 시작된 연구를 계속하자. 그곳에서 도입된 표기법을 사용하여 $\mu = E(\xi_i^m) > 1$이며 $\mathrm{var}(\xi_i^m) = \sigma^2 < \infty$라 가정하자. $X_n = Z_n/\mu^n$이라 하자. Theorem 4.4.8에서 $m = n-1$로 취하고 재정비하면 다음을 얻는다.

$$E(X_n^2|\mathcal{F}_{n-1}) = X_{n-1}^2 + E((X_n - X_{n-1})^2|\mathcal{F}_{n-1})$$

둘째 항을 계산하기 위해 다음이 성립함을 관찰하자.

$$\begin{aligned}
E((X_n - X_{n-1})^2|\mathcal{F}_{n-1}) &= E((Z_n/\mu^n - Z_{n-1}/\mu^{n-1})^2|\mathcal{F}_{n-1}) \\
&= \mu^{-2n}E((Z_n - \mu Z_{n-1})^2|\mathcal{F}_{n-1})
\end{aligned}$$

Exercise 4.1.2로부터 $\{Z_{n-1} = k\}$에서 다음이 성립함이 따라온다.

$$E((Z_n - \mu Z_{n-1})^2|\mathcal{F}_{n-1}) = E\left(\left(\sum_{i=1}^{k} \xi_i^n - \mu k\right)^2 \middle| \mathcal{F}_{n-1}\right) = k\sigma^2 = Z_{n-1}\sigma^2$$

위 세 식들을 조합하면 ($E(Z_{n-1}/\mu^{n-1}) = EZ_0 = 1$이므로) 다음을 얻는다.

$$EX_n^2 = EX_{n-1}^2 + E(Z_{n-1}\sigma^2/\mu^{2n}) = EX_{n-1}^2 + \sigma^2/\mu^{n+1}$$

이제 $EX_0^2 = 1$이므로 $EX_1^2 = 1 + \sigma^2/\mu^2$이며 따라서 귀납법에 의해 다음이 성립한다.

$$EX_n^2 = 1 + \sigma^2 \sum_{k=2}^{n+1} \mu^{-k}$$

이는 $\sup EX_n^2 < \infty$이며 따라서 L^2에서 $X_n \to X$이고 그러므로 $EX_n \to EX$임을 보여준다. 모든 n에 대하여 $EX_n = 1$이며 따라서 $EX = 1$이고 그러므로 X는 항등적으로 0이지 않다. Exercise 4.3.11로부터 $\{X > 0\} = \{\forall n\ Z_n > 0\}$임이 따라온다.

4.5 Square Integrable Martingales (제곱 적분가능 마팅게일)*

4.6 Uniform Integrability, Convergence in L^1 (균등적분가능성, L^1 수렴성)

이 절에서 우리는 마팅게일이 L^1에서 수렴할 필요충분조건을 제시하겠다. 핵심은 다음 정의이다: 확률변수들의 족 X_i $(i \in I)$가 **균등적분가능(uniformly integrable)**이라는 것의 정의는 다음을 만족시키는 것이다.

$$\lim_{M\to\infty}\left(\sup_{i\in I} E(|X_i|; |X_i| > M)\right) = 0$$

M을 충분히 크게 선택하여 위 상한이 1 미만이도록 할 경우 다음이 성립함이 따라온다.

$$\sup_{i \in I} E|X_i| \leq M + 1 < \infty$$

이 언급은 나중에 여러 번 유용하게 사용될 것이다.

균등적분가능 족의 자명한 예시는 적분가능 확률변수에 의해 지배되는 확률변수들의 족이다. i.e. 어떠한 Y가 존재하여 $|X_i| \leq Y$이며 $EY < \infty$이다. 우리의 첫 번째 결과는 균등적분가능 족이 매우 클 수 있음을 보여주는 흥미로운 예시를 제공한다.

Theorem 4.6.1. 주어진 확률공간 $(\Omega, \mathcal{F}_0, P)$와 $X \in L^1$에 대하여 $\{E(X|\mathcal{F}) : \mathcal{F} \subset \mathcal{F}_0$가 σ-체이다$\}$는 균등적분가능하다.

Proof. 만약 A_n이 $P(A_n) \to 0$을 만족시키는 집합열이면 지배 수렴 정리에 의해 $E(|X|; A_n) \to 0$이다. 마지막 결과로부터 $\varepsilon > 0$이 주어진 경우 $\delta > 0$을 선택하여 $P(A) \leq \delta$이면 $E(|X|; A) \leq \varepsilon$를 만족시키도록 할 수 있음이 따라온다. (만약 그렇지 않다면 집합 A_n들이 존재하여 $P(A_n) \leq 1/n$이며 $E(|X|; A_n) > \varepsilon$이고 이는 모순이다.)

M을 충분히 크게 선택하여 $E|X|/M \leq \delta$를 만족시키도록 하자. $\{E(|X||\mathcal{F}) > M\} \in \mathcal{F}$이므로 Jensen 부등식과 조건부기댓값의 정의는 다음을 함의한다.

$$\begin{aligned} E(|E(X|\mathcal{F})|; |E(X|\mathcal{F})| > M) &\leq E(E(|X||\mathcal{F}); E(|X||\mathcal{F}) > M) \\ &= E(|X|; E(|X||\mathcal{F}) > M) \end{aligned}$$

Chebyshev 부등식을 사용하고 M의 정의를 상기하면 다음을 얻는다.

$$P\{E(|X||\mathcal{F}) > M\} \leq E\{E(|X||\mathcal{F})\}/M = E|X|/M \leq \delta$$

그러므로 δ의 선택에 의해 다음이 성립한다.

$$E(|E(X|\mathcal{F})|; |E(X|\mathcal{F})| > M) \leq \varepsilon \qquad (\forall \mathcal{F})$$

ε가 임의로 선택되었으므로 이러한 족은 균등적분가능하다. □

균등적분가능성을 확인하는 일반적인 방법은 다음을 사용하는 것이다:

Theorem 4.6.2. $\varphi \geq 0$이 $x \to \infty$에서 $\varphi(x)/x \to \infty$를 만족시키는 임의의 함수라 하자. e.g. $\varphi(x) = x^p$ $(p > 1)$ 또는 $\varphi(x) = x \log^+ x$. 만약 모든 $i \in I$에 대하여 $E\varphi(|X_i|) \leq C$이면 $\{X_i : i \in I\}$가 균등적분가능하다.

Proof. $\varepsilon_M = \sup\{x/\varphi(x) : x \geq M\}$이라 하자. $i \in I$에 대하여 다음이 성립한다.

$$E(|X_i|; |X_i| > M) \leq \varepsilon_M E(\phi(|X_i|); |X_i| > M) \leq C\varepsilon_M$$

또한 $M \to \infty$에서 $\varepsilon_M \to 0$이다. □

균등적분가능성과 L^1 수렴성 간의 관계는 다음에 의해 설명된다:

Theorem 4.6.3. 모든 n에 대하여 $E|X_n| < \infty$라 하자. 만약 $X_n \to X$로 확률에서 수렴하면 다음이 동치이다:

(i) $\{X_n : n \geq 0\}$이 균등적분가능하다.
(ii) $X_n \to X$로 L^1 수렴한다.
(iii) $E|X_n| \to E|X| < \infty$

Proof. (i) $\Rightarrow$ (ii). 다음과 같다 하자.

$$\varphi_M(x) = \begin{cases} M & (x \geq M) \\ x & (|x| \leq M) \\ -M & (x \leq -M) \end{cases}$$

삼각부등식은 다음을 함의한다.

$$|X_n - X| \leq |X_n - \varphi_M(X_n)| + |\varphi_M(X_n) - \varphi_M(X)| + |\varphi_M(X) - X|$$

$|\varphi_M(Y) - Y| = (|Y| - M)^+ \le |Y| 1_{(|Y|>M)}$이므로 기댓값을 취하면 다음을 얻는다.

$$E|X_n - X| \le E|\varphi_M(X_n) - \varphi_M(X)| + E(|X_n|; |X_n| > M) + E(|X|; |X| > M)$$

Theorem 2.3.4는 $\varphi_M(X_n) \to \varphi_M(X)$로 확률에서 수렴함을 함의하며, 따라서 유계 수렴 정리에 의해 첫째 항이 0으로 수렴한다. (Exercise 2.3.5를 참조하라.) 만약 $\varepsilon > 0$이며 M이 충분히 크면 균등적분가능성은 둘째 항이 ε 이하임을 함의한다. 셋째 항의 상계를 설정하기 위해 균등적분가능성이 $\sup E|X_n| < \infty$임을 함의함을 상기하라. 따라서 (Exercise 2.3.4에 주어진 형태의) Fatou 보조정리는 $E|X| < \infty$임을 함의한다. 필요하다면 M을 더 크게 설정하는 것으로 셋째 항이 ε 이하이도록 할 수 있다. 이러한 사실들을 조합하면 $\limsup E|X_n - X| \le 2\varepsilon$임을 알 수 있다. ε가 임의로 선택되었으므로 이는 (ii)를 증명한다.
(ii) $\Rightarrow$ (iii). Jensen 부등식은 다음을 함의한다.

$$|E|X_n| - E|X|| \le E||X_n| - |X|| \le E|X_n - X| \to 0$$

(iii) $\Rightarrow$ (i). 다음과 같다 하자.

$$\psi_M(x) = \begin{cases} x & (x \in [0, M-1]) \\ 0 & (x \in [M, \infty)) \\ \text{선형} & (x \in [M-1, M]) \end{cases}$$

지배 수렴 정리는 만약 M이 크다면 $E|X| - E\psi_M(|X|) < \varepsilon/2$임을 함의한다. 증명의 첫째 부분에서와 같이 유계 수렴 정리는 $E\psi_M(|X_n|) \to E\psi_M(|X|)$임을 함의한다. 그러므로 (iii)을 사용하면 $n \ge n_0$인 경우 다음이 성립함을 알 수 있다.

$$\begin{aligned} E(|X_n|; |X_n| > M) &\le E|X_n| - E\psi_M(|X_n|) \\ &\le E|X| - E\psi_M(|X|) + \varepsilon/2 < \varepsilon \end{aligned}$$

M을 더 크게 선택하는 것으로 $0 \le n < n_0$에 대해서도 $E(|X_n|; |X_n| > M) \le \varepsilon$이도록 할 수 있으며 따라서 X_n이 균등적분가능하다. □

우리는 이제 이 절의 주요 정리들을 기술할 준비가 되었다. 우리는 이미 작업의 대부분을 수행했으므로 증명이 짧을 것이다.

Theorem 4.6.4. 열마팅게일에 대하여 다음이 동치이다:
(i) 이것이 균등적분가능하다.
(ii) 이것이 a.s. 수렴하며 L^1 수렴한다.
(iii) 이것이 L^1 수렴한다.

Proof. (i) $\Rightarrow$ (ii). 균등적분가능성은 $\sup E|X_n| < \infty$임을 함의하며 따라서 마팅게일 수렴 정리는 $X_n \to X$ a.s.임을 함의한다. 또한 Theorem 4.6.3은 $X_n \to X$로 L^1 수렴함을 함의한다.
(ii) $\Rightarrow$ (iii). 자명. (iii) $\Rightarrow$ (i). $X_n \to X$로 L^1 수렴함은 $X_n \to X$로 확률에서 수렴함을 함의하며 (Lemma 2.2.2를 참조하라) 따라서 이는 Theorem 4.6.3에서 따라온다. □

마팅게일에 대한 Theorem 4.6.4의 유사품을 증명하기 전에 우리는 나중에 유용하게 사용될 논의의 두 가지 부분을 분리하겠다.

Lemma 4.6.5. 만약 적분가능 확률변수 X_n들이 $X_n \to X$로 L^1 수렴하면 다음이 성립한다.

$$E(X_n; A) \to E(X; A)$$

Proof. $|EX_m 1_A - EX 1_A| \le E|X_m 1_A - X 1_A| \le E|X_m - X| \to 0$ □

Lemma 4.6.6. 만약 마팅게일 X_n이 $X_n \to X$로 L^1 수렴하면 $X_n = E(X|\mathcal{F}_n)$이다.

Proof. 마팅게일 성질은 $m > n$이면 $E(X_m|\mathcal{F}_n) = \mathcal{F}_n$임을 함의하며 따라서 만약 $A \in \mathcal{F}_n$이면 $E(X_n; A) = E(X_m; A)$를 만족시킨다. Lemma 4.6.5는 $E(X_m; A) \to E(X; A)$임을 함의하며 따라서 모든 $A \in \mathcal{F}_n$에 대하여 $E(X_n; A) = E(X; A)$이다. 조건부기댓값의 정의를 상기하면 $X_n = E(X|\mathcal{F}_n)$임이 따라온다. □

Theorem 4.6.7. 마팅게일에 대하여 다음이 동치이다:
(i) 이것이 균등적분가능하다.
(ii) 이것이 a.s. 수렴하며 L^1 수렴한다.

(iii) 이것이 L^1 수렴한다.
(iv) 적분 가능 확률변수 X가 존재하여 $X_n = E(X|\mathcal{F}_n)$을 만족시킨다.

Proof. (i) ⇒ (ii). 마팅게일이 열마팅게일이므로 이는 Theorem 4.6.4에서 따라온다. (ii) ⇒ (iii). 자명. (iii) ⇒ (iv). 이는 Lemma 4.6.6에써 따라온다. (iv) ⇒ (i). 이는 Theorem 4.6.1에서 따라온다. □

다음 결과는 Lemma 4.6.6과 관련이 있지만 다른 방향으로 진행한다.

Theorem 4.6.8. $\mathcal{F}_n \uparrow \mathcal{F}_\infty$라 가정하자. i.e. $\mathcal{F}_n$이 σ-체들의 증가 열이며 $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$이다. 그 경우 $n \to \infty$에서 다음이 성립한다.

$$E(X|\mathcal{F}_n) \to E(X|\mathcal{F}_\infty) \quad \text{로 a.s. 수렴 및 } L^1 \text{ 수렴}$$

Proof. 첫째 단계는 $m > n$이면 Theorem 4.1.13에 의해 다음이 성립함을 기억하는 것이다.

$$E(E(X|\mathcal{F}_m)|\mathcal{F}_n) = E(X|\mathcal{F}_n)$$

그러므로 $Y_n = E(X|\mathcal{F}_n)$이 마팅게일이다. Theorem 4.6.1은 Y_n이 균등적분가능함을 함의하며 따라서 Theorem 4.6.7은 Y_n이 극한 Y_∞로 a.s. 수렴하며 L^1 수렴함을 함의한다. Y_n의 정의와 Lemma 4.6.6은 $E(X|\mathcal{F}_n) = Y_n = E(Y_\infty|\mathcal{F}_n)$임을 하므이하며 따라서 다음이 성립한다.

$$\int_A X\, dP = \int_A Y_\infty\, dP \qquad (\forall A \in \mathcal{F}_n)$$

X와 Y_∞가 적분 가능하며 $\bigcup_n \mathcal{F}_n$이 π-계이므로 π-λ 정리는 마지막 결과가 모든 $A \in \mathcal{F}_\infty$에 대하여 성립함을 함의한다. $Y_\infty \in \mathcal{F}_\infty$이므로 $Y_\infty = E(X|\mathcal{F}_\infty)$임이 따라온다. □

Theorem 4.6.8의 즉각적인 따름정리는 다음과 같다:

Theorem 4.6.9 (Lévy's 0-1 law (L'evy 0-1 법칙))**.** 만약 $\mathcal{F}_n \uparrow \mathcal{F}_\infty$이며 $A \in \mathcal{F}_\infty$이면 $E(1_A|\mathcal{F}_n) \to 1_A$ a.s. 이다.

Chung의 말을 인용하면: "독자는 이 결과의 의미에 대하여 숙고하고 그것이 자명한지 놀라운지를 스스로 판단해야 한다." 우리는 이제 두 가지 관점에 대하여 논의하겠다.

"이는 자명하다." $1_A \in \mathcal{F}_\infty$이며 $\mathcal{F}_n \uparrow \mathcal{F}_\infty$이므로 $\mathcal{F}_n$에서 주어진 정부에 대한 1_A의 최적 추측은 1_A($\mathcal{F}_\infty$에서 주어진 최적 추측)에 접근해야 한다.

"이는 놀랍다." $X_1, X_2, \ldots$가 독립이며 $A \in \mathcal{T}$ (꼬리 σ-체)라 하자. 각각의 n에 대하여 A는 $\mathcal{F}_n$에 독립적이며 따라서 $E(1_A|\mathcal{F}_n) = P(A)$이다. $n \to \infty$에서 좌변은 1_A로 a.s. 수렴하며 따라서 $P(A) = 1_A$ a.s.이다. $P(A) \in \{0, 1\}$임이 따라온다. i.e. 우리는 Kolmogorov 0-1 법칙을 증명했다.

마지막 논의는 Theorem 4.6.9가 '참이 되기에는 너무 특이하거나 희한함'을 보여주지는 않음을 기억해 두라. 그러나 이것과 Theorem 4.6.9의 다른 응용은 이것이 매우 유용한 결과임을 보여준다.

Theorem 4.6.8의 더 기술적인 결과로 다음이 있다:

Theorem 4.6.10 (Dominated convergence theorem for conditional expectation (조건부기댓값에 대한 지배 수렴 정리))**.** $Y_n \to Y$ a.s.이며 모든 n에 대하여 $|Y_n| \le Z$이고 $EZ < \infty$라 하자. 만약 $\mathcal{F}_n \uparrow \mathcal{F}_\infty$이면 다음이 성립한다.

$$E(Y_n|\mathcal{F}_n) \to E(Y|\mathcal{F}_\infty) \quad \text{a.s.}$$

Proof. $W_N = \sup\{|Y_n - Y_m| : n, m \ge N\}$이라 하자. $W_N \le 2Z$이므로 $EW_N < \infty$이다. 단조성 (4.1.2)를 사용하고 Theorem 4.6.8을 W_N에 적용하면 다음을 얻는다.

$$\limsup_{n\to\infty} E(|Y_n - Y||\mathcal{F}_n) \le \lim_{n\to\infty} E(W_N|\mathcal{F}_n) = E(W_N|\mathcal{F}_\infty)$$

위 결과는 모든 N에 대하여 참이며 $N \uparrow \infty$에서 $W_N \downarrow 0$이므로 (4.1.3)은 $E(W_N|\mathcal{F}_\infty) \downarrow 0$임을 함의한다. Jensen 부등식은 다음을 제공한다.

$$|E(Y_n|\mathcal{F}_n) - E(Y|\mathcal{F}_n)| \le E(|Y_n - Y||\mathcal{F}_n) \to 0 \qquad (n \to \infty\text{에서 a.s.})$$

Theorem 4.6.8은 $E(Y|\mathcal{F}_n) \to E(Y|\mathcal{F}_\infty)$ a.s.임을 함의한다. 요구된 결과는 위 두 결과와 삼각부등식에서 따라온다. □

Example 4.6.11. $X_1, X_2, \ldots$가 균등적분가능하며 X로 a.s. 수렴한다 하자. Theorem 4.6.3은 $X_n \to X$로 L^1 수렴함을 함의한다. 이를 Exercise 4.6.7과 조합하면 $E(X_n|\mathcal{F}) \to E(X|\mathcal{F})$로 L^1 수렴함을 보일 수 있다. 우리는 이제 $E(X_n|\mathcal{F})$가 a.s. 수렴할 필요가 없음을 보이겠다. $Y_1, Y_2, \ldots$와 $Z_1, Z_2, \ldots$가 다음을 만족시키는 독립 확률변수들이라 하자.

$$P(Y_n = 1) = 1/n \quad P(Y_n = 0) = 1 - 1/n$$
$$P(Z_n = n) = 1/n \quad P(Z_n = 0) = 1 - 1/n$$

$X_n = Y_n Z_n$이라 하자. $P(X_n > 0) = 1/n^2$이며 따라서 Borel-Cantelli 보조정리는 $X_n \to 0$ a.s.임을 함의한다. $E(X_n; |X_n| \geq 1) = n/n^2$이며 따라서 X_n은 균등적분가능하다. $\mathcal{F} = \sigma(Y_1, Y_2, \ldots)$라 하자.

$$E(X_n||mcF) = Y_n E(Z_n|\mathcal{F}) = Y_n EZ_n = Y_n$$

$Y_n \to 0$으로 L^1 수렴하지만 a.s. 수렴하지는 않으므로 같은 것이 $E(X_n|\mathcal{F})$에 대해서도 참이다.

4.7 Backwards Martingales (후진 마팅게일)

후진 마팅게일(backwards martingale)(어떠한 저자들은 **반전 마팅게일(reversed martingale)**이라 한다)은 음의 정수에 의해 첨자화된 마팅게일이다. i.e. σ-체들의 증가 열 $\mathcal{F}_n$에 적응된 X_n $(n \leq 0)$ 중 다음을 만족시키는 것이다.

$$E(X_{n+1}|\mathcal{F}_n) = X_n \qquad (n \leq -1)$$

$n \downarrow -\infty$에서 σ-체들이 감소하므로 후진 마팅게일에 대한 수렴 이론은 특별히 간단하다.

Theorem 4.7.1. $X_{-\infty} = \lim_{n \to -\infty} X_n$이 a.s. 존재하며 동시에 L^1 극한이다.

Proof. U_n이 $[a, b]$에서의 $X_{-n}, \ldots, X_0$에 의한 상승교차의 수라 하자. 상승교차 부등식(Theorem 4.2.10)은 $(b-a)EU_n \leq E(X_0 - a)^+$임을 함의한다. $n \to \infty$라 하고 단조 수렴 정리를 사용하면 $EU_\infty < \infty$임을 얻는다. 따라서 Theorem 4.2.11의 증명 이후의 언급에 의해 극한이 a.s. 존재한다. 마팅게일 성질은 $X_n = E(X_0|\mathcal{F}_n)$임을 함의하며 따라서 Theorem 4.6.1은 X_n이 균등적분가능함을 함의하고 Theorem 4.6.3은 수렴이 L^1 내에서 발생함을 말해준다. □

다음 결과는 Theorem 4.7.1의 극한을 찾아 준다.

Theorem 4.7.2. 만약 $X_{-\infty} = \lim_{n \to -\infty} X_n$이며 $\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_n$이면 $X_{-\infty} = E(X_0|\mathcal{F}_{-\infty})$이다.

Proof. 명백히 $X_{-\infty} \in \mathcal{F}_{-\infty}$이다. $X_n = E(X_0|\mathcal{F}_n)$이므로 만약 $A \in \mathcal{F}_{-\infty} \subset \mathcal{F}_n$이면 다음이 성립한다.

$$\int_A X_n \, dP = \int_A X_0 \, dP$$

Theorem 4.7.1과 Lemma 4.6.5는 $E(X_n; A) \to E(X_{-\infty}; A)$임을 함의하며 따라서 모든 $A \in \mathcal{F}_{-\infty}$에 대하여 다음이 성립한다.

$$\int_A X_{-\infty} \, dP = \int_A X_0 \, dP$$

이는 요구된 결과를 증명한다. □

다음 결과는 후진 Theorem 4.6.8이다.

Theorem 4.7.3. 만약 $n \downarrow -\infty$에서 $\mathcal{F}_n \downarrow \mathcal{F}_{-\infty}$이면 (i.e. $\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_n$이면) 다음이 성립한다.

$$E(Y|\mathcal{F}_n) \to E(Y|\mathcal{F}_{-\infty}) \qquad \text{(a.s. 수렴하며 } L^1 \text{ 수렴)}$$

Proof. $X_n = E(Y|\mathcal{F}_n)$이 후진 마팅게일이다. 따라서 Theorem 4.7.1과 4.7.2는 $n \downarrow -\infty$에서 X_n이 다음과 같은 $X_{-\infty}$로 a.s. 수렴하며 L^1 수렴함을 함의한다.

$$X_{-\infty} = E(X_0|\mathcal{F}_{-\infty}) = E(E(Y|\mathcal{F}_0)|\mathcal{F}_{-\infty}) = E(Y|\mathcal{F}_{-\infty})$$

□

후진 마팅게일에 대한 수렴 이론이 간단함에도 불구하고 좋은 응용이 존재한다. 이 절의 남은 부분에서는 Section 4.1에서 활용했던 특수한 공간으로 돌아갈 것이며 따라서 그곳에서 주어진 정의들을 사용할 것이다. 즉 다음과 같다 하자.

$$\begin{aligned}
&\Omega = \{(\omega_1, \omega_2, \ldots) : \omega_i \in S\} \\
&\mathcal{F} = \mathcal{S} \times \mathcal{S} \times \cdots \\
&X_n(\omega) = \omega_n
\end{aligned}$$

$\mathcal{E}_n$이 $n+1, n+2, \ldots$를 고정하는 치환 하에서 불변인 사건들에 의해 생성된 σ-체라 하고 $\mathcal{E} = \bigcap_n \mathcal{E}_n$이 교환가능 σ-체라 하자.

Example 4.7.4 (Strong law of large numbers (대수의 강법칙))**.** $\xi_1, \xi_2, \ldots$가 i.i.d.이며 $E|\xi_i| < \infty$를 만족시킨다 하자. $S_n = \xi_1 + \cdots + \xi_n$이며 $X_{-n} = S_n/n$이라 하자. 다음과 같다 하자.

$$\mathcal{F}_{-n} = \sigma(S_n, S_{n+1}, S_{n+2}, \ldots) = \sigma(S_n, \xi_{n+1}, \xi_{n+2}, \ldots)$$

$E(X_{-n}|\mathcal{F}_{-n-1})$을 계산하기 위해 만약 $j, k \le n+1$이면 대칭성에 의해 $E(\xi_j|\mathcal{F}_{-n-1}) = E(\xi_k|\mathcal{F}_{-n-1})$이며 따라서 다음이 성립함을 기억해 두라.

$$\begin{aligned}
E(\xi_{n+1}|\mathcal{F}_{-n-1}) &= \frac{1}{n+1}\sum_{k=1}^{n+1} E(\xi_k|\mathcal{F}_{-n-1}) \\
&= \frac{1}{n+1} E(S_{n+1}|\mathcal{F}_{-n-1}) = \frac{S_{n+1}}{n+1}
\end{aligned}$$

$X_{-n} = (S_{n+1} - \xi_{n+1})/n$이므로 다음이 따라온다.

$$\begin{aligned}
E(X_{-n}|\mathcal{F}_{-n-1}) &= E(S_{n+1}/n|\mathcal{F}_{-n-1}) - E(\xi_{n+1}/n|\mathcal{F}_{-n-1}) \\
&= \frac{S_{n+1}}{n} - \frac{S_{n+1}}{n(n+1)} = \frac{S_{n+1}}{n+1} = X_{-n-1}
\end{aligned}$$

마지막 계산은 X_{-n}이 후진 마팅게일임을 보여주며 따라서 Theorem 4.7.1, 4.7.2로부터 $\lim_{n\to\infty} S_n/n = E(X_{-1}|\mathcal{F}_{-\infty})$임이 따라온다. $\mathcal{F}_{-n} \subset \mathcal{E}_n$이므로 $\mathcal{F}_{-\infty} \subset \mathcal{E}$이다. Hewitt-Savage 0-1 법칙(Theorem 2.5.4)은 $\mathcal{E}$가 자명함을 말해주며, 따라서 다음이 성립한다.

$$\lim_{n\to\infty} S_n/n = E(X_{-1}) \quad \text{a.s.}$$

(중간 생략)

Example 4.7.8 (de Finetti's Theorem (de Finetti 정리))**.** 열 $X_1, X_2, \ldots$가 **교환가능(exchangeable)** 이라는 것의 정의는 각각의 n과 $\{1, \ldots, n\}$의 치환 π에 대하여 $(X_1, \ldots, X_n)$과 $(X_{\pi(1)}, \ldots, X_{\pi(n)})$이 동일한 분포를 가지는 것이다.

Theorem 4.7.9 (de Finetti's Theorem (de Finetti 정리))**.** 만약 $X_1, X_2, \ldots$가 교환가능하면 $X_1, X_2, \ldots$ 들의 $\mathcal{E}$에서의 조건부기댓값들이 독립동분포이다.

Proof. Lemma 4.7.7의 증명에서의 표기법을 도입하여 그곳의 첫 번째 계산을 반복하면 임의의 교환 가능 열에 대하여 (합의 모든 항들이 서로 같으므로) 다음이 성립함을 알 수 있다:

$$\begin{aligned}
A_n(\varphi) = E(A_n(\varphi)|\mathcal{E}_n) &= \frac{1}{(n)_k}\sum_i E(\varphi(X_{i_1}, \ldots, X_{i_k})|\mathcal{E}_n) \\
&= E(\varphi(X_1, \ldots, X_k)|\mathcal{E}_n)
\end{aligned}$$

다시 Theorem 4.7.3은 다음을 함의한다.

$$A_n(\varphi) \to E(\varphi(X_1, \ldots, X_k)|\mathcal{E}) \tag{4.7.1}$$

그러나 이 경우 $\mathcal{E}$가 비자명할 수 있으므로 극한이 $E(\varphi(X_1, \ldots, X_k))$임을 보일 수 있을 것이라 기대해서는 안 된다.

f와 g가 각각 $\mathbb{R}^{k-1}$과 $\mathbb{R}$에서의 유계 함수라 하자. 만약 $I_{n,k}$가 서로 다른 정수 $1 \le i_1, \ldots, i_k \le n$들의 모든 열들의 집합이라 하면 다음이 성립한다.

$$(n)_{k-1} A_n(f) A_n(g) = \sum_{i \in I_{n,k-1}} f(X_{i_1}, \ldots, X_{i_{k-1}}) \sum_m g(X_m)$$
$$= \sum_{i \in I_{n,k}} f(X_{i_1}, \ldots, X_{i_{k-1}}) g(X_{i_k}) + \sum_{i \in I_{n,k-1}} \sum_{j=1}^{k-1} f(X_{i_1}, \ldots, X_{i_{k-1}}) g(X_{i_j})$$

$\varphi(x_1, \ldots, x_k) = f(x_1, \ldots, x_{k-1}) g(x_k)$라 설정하고 다음 식

$$\frac{(n)_{k-1} n}{(n)_k} = \frac{n}{(n-k+1)} \quad , \quad \frac{(n)_{k-1}}{(n)_k} = \frac{1}{(n-k+1)}$$

을 적용한 후 정리하면 다음을 얻는다.

$$A_n(\varphi) = \frac{n}{n-k+1} A_n(f) A_n(g) - \frac{1}{n-k+1} \sum_{j=1}^{k-1} A_n(\varphi_j)$$

여기에서 $\varphi_j(x_1, \ldots, x_{k-1}) = f(x_1, \ldots, x_{k-1}) g(x_j)$이다. φ, f, g와 모든 φ_j에 (4.7.2)를 적용하면 다음을 얻는다.

$$E(f(X_1, \ldots, X_{k-1}) g(X_k) | \mathcal{E}) = E(f(X_1, \ldots, X_{k-1}) | \mathcal{E}) E(g(X_k) | \mathcal{E})$$

귀납법으로부터 다음이 따라온다.

$$E\left(\prod_{j=1}^k f_j(X_j) \middle| \mathcal{E} \right) = \prod_{j=1}^k E(f_j(X_j) | \mathcal{E})$$

□

(생략)

4.8 Optimal Stopping Problem (최적 정지 문제)

이 절에서 우리는 만약 X_n이 열마팅게일이며 $M \le N$이 정지시간들이면 $EX_M \le EX_N$이라 결론지을 수 있게 해 주는 다수의 결과들을 증명할 것이다. Example 4.2.13은 이것이 항상 참이지는 않음을 보여주지만 Exercise 4.4.2는 N이 유계인 경우 이것이 참임을 보여주며 따라서 우리는 비유계 N의 경우에 집중할 것이다.

Theorem 4.8.1. 만약 X_n이 균등적분가능 열마팅게일이면 임의의 정지시간 N에 대하여 $X_{N \wedge n}$이 균등적분가능하다.

Theorem 4.2.5에서와 같이 위 결과는 우마팅게일에 대하여 $\ge$, 마팅게일에 대하여 $=$를 가지고 성립한다. 이는 다음 두 정리들에 대해서도 마찬가지이다.

Proof. X_n^+이 열마팅게일이며 따라서 Theorem 4.4.1은 $EXk_{N \wedge n} \le EX_n^+$임을 함의한다. X_n^+이 균등적분가능하므로 정의 이후의 언급에서 다음이 따라온다.

$$\sup_n EX_{N \wedge n}^+ \le \sup_n EX_n^+ < \infty$$

이제 마팅게일 수렴 정리(Theorem 4.2.11)는 $X_{N \wedge n} \to X_N$ a.s.이며 $E|X_N| < \infty$임을 보여준다. (여기에서 $X_\infty = \lim_n X_n$이다.) 이제 남은 부분은 간단하다. 다음과 같이 표현하자.

$$E(|X_{N \wedge n}|; |X_{N \wedge n}| > K) = E(|X_N|; |X_N| > K, N \le n) + E(|X_n|; |X_n| > K, N > n)$$

$E|X_N < \infty|$이며 X_n이 균등 적분 가능하므로 만약 K가 크다면 각각의 항이 $\varepsilon/2$ 미만이다. □

Theorem 4.8.1의 증명의 마지막 계산에서 다음을 얻는다:

Theorem 4.8.2. 만약 $E|X_N| < \infty$이며 $X_n 1_{(N>n)}$이 균등적분가능하면 $X_{N\wedge n}$이 균등적분가능하며 따라서 $EX_0 \le EX_N$이다.

Theorem 4.8.3. 만약 X_n이 균등적분가능 열마팅게일이면 임의의 정지시간 $N \le \infty$에 대하여 $EX_0 \le EX_N \le EX_\infty$이다. (여기에서 $X_\infty = \lim X_n$이다.)

Proof. Theorem 4.4.1은 $EX_0 \le EX_{N\wedge n} \le EX_n$임을 함의한다. $n \to \infty$의 극한을 취하고 Theorem 4.8.1, 4.6.4를 적용하면 $X_{N\wedge n} \to X_N, X_n \to X_\infty$로 L^1 수렴하며 이는 요구된 결과를 제공한다. □

다음 결과는 균등적분가능성을 필요로 하지 않는다.

Theorem 4.8.4. 만약 X_n이 음 아닌 우마팅게일이며 $N \le \infty$가 정지시간이면 $EX_0 \ge EX_N$이다. (여기에서 $X_\infty = \lim X_n$이며, Theorem 4.2.12에 의해 이것이 존재한다.)

Proof. Theorem 4.4.1과 Fatou 보조정리에 의해 다음이 성립한다.

$$EX_0 \ge \liminf_{n\to\infty} EX_{N\wedge n} \ge EX_N$$

□

다음 결과는 몇몇 상황에서 유용하다.

Theorem 4.8.5. X_n이 열마팅게일이며 $E(|X_{n+1} - X_n||\mathcal{F}_n) \le B$ a.s.라 하자. 만약 N이 정지시간이며 $EN < \infty$이면 $X_{N\wedge n}$은 균등적분가능하며 따라서 $EX_N \ge EX_0$이다.

Proof. 다음을 관찰하는 것으로 시작하자.

$$|X_{N\wedge n}| \le |X_0| + \sum_{m=0}^{\infty} |X_{m+1} - X_m| 1_{(N>m)}$$

균등적분가능성을 증명하기 위해서는 우변이 유한 기댓값을 가짐을 보여면 충분하다; 그 경우 $|X_{N\wedge n}|$이 적분가능 확률변수에 의해 지배된다. 이제 $\{N > m\} \in \mathcal{F}_m$이므로 다음이 성립한다.

$$E(|X_{m+1} - X_m|; N > m) = E(E(|X_{m+1} - X_m||\mathcal{F}_m); N > m) \le BP(N > m)$$

또한 $E\sum_{m=0}^{\infty} |X_{m+1} - X_m| 1_{(N>m)} \le B\sum_{m=0}^{\infty} P(N > m) = BEN < \infty$이다. □

4.8.1 Applications to Random Walks (난보에의 응용)

$\xi_1, \xi_2, \ldots$가 i.i.d.이고 S_0가 상수이며 $S_n = S_0 + \xi_1 + \cdots + \xi_n$이라 하자. 또한 $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$이라 하자. Section 4.2의 3가지 마팅게일을 이용하여 몇 가지 결과들을 유도할 것이다.

Linear martingale (선형마팅게일). 만약 $\mu = E\xi_i$라 하면 $X_n = S_n - n\mu$가 마팅게일이다. (Example 4.2.1을 참조하라.)

선형마팅게일에 Theorem 4.8.5를 적용하면 다음을 얻는다.

Theorem 4.8.6 (Wald's equation (Wald 방정식))**.** 만약 $\xi_1, \xi_2, \ldots$가 i.i.d.이며 $E\xi_i = \mu$를 만족시키고 $S_n = \xi_1 + \cdots + \xi_n$이며 N이 $EN < \infty$를 만족시키는 정지시각이면 $ES_N = \mu EN$이다.

Proof. $X_n = S_n - n\mu$라 하고 $E(|X_{n+1} - X_n||\mathcal{F}_n) = E|\xi_i - \mu|$임을 기억해 두라. □

Quadratic martingale (2차마팅게일). $E\xi_i = 0$이며 $E\xi_i^2 = \sigma^2 \in (0, \infty)$라 하자. 그 경우 $X_n = S_n - n\sigma^2$가 마팅게일이다. (Example 4.2.2를 참조하라.)

Exponential martingale (지수마팅게일). $\phi(\theta) = E\exp(\theta\xi_i) < \infty$라 하자. 그 경우 $X_n = \exp(\theta S_n)/\phi(\theta)^n$이 마팅게일이다. (Example 4.2.3을 참조하라.)

Theorem 4.8.7 (Symmetric simple random walk (대칭단순난보)). 대칭단순난보는 $P(\xi_i = 1) = P(\xi_i = -1) = 1/2$인 특수한 경우를 나타낸다. $S_0 = x$이며 $N = \min\{n : S_n \in (a, b)\}$라 하자. 시작점을 나타내기 위해 아래첨자 x를 표기하자. 다음이 성립한다.

(a)
$$P_x(S_N = a) = \frac{b-x}{b-a} \quad , \quad P_x(S_N = b) = \frac{x-a}{b-a}$$

(b) $E_0 N = -ab$이며 따라서 $E_x N = (b-x)(a-x)$이다.

$T_x = \min\{n : S_n = x\}$라 하자. $a = 0, x = 1, b = M$으로 선택하면 다음을 얻는다.

$$P_1(T_M < T_0) = \frac{1}{M} \quad , \quad P1(T_M \geq T_0) = \frac{M-1}{M}$$

첫째는 (4.4.1)을 증명한다. 둘째에서 $M \to \infty$라 하면 $P_1(T_0 < \infty) = 1$임을 얻는다.

Proof. (a) $P(N < \infty) = 1$임을 보이기 위해 크기가 $+1$되는 연이은 $(b-a)$회의 단계가 존재한다면 구간을 벗어남을 기억해 두라. 이로부터 다음이 따라온다.

$$P(N > m(b-a)) \leq (1 - 2^{-(b-a)})^m$$

따라서 $EN < \infty$이다.

명백히 $E|S_N| < \infty$이며 $S_n 1_{\{N>n\}}$이 균등적분가능하므로 Theorem 4.8.2를 사용하면 다음을 얻는다.

$$x = ES_N = aP_x(S_N = a) + b[1 - P_x(S_N = a)]$$

정리하면 $P_x(S_N = a) = (b-x)/(b-a)$이다. 1에서 이를 빼면 $P_x(S_N = b) = (x-a)/(b-a)$이다.
(b) 둘째 결과는 첫째 결과에서 즉시 따라온다.

유계 종료시간 $N \wedge n$에 대하여 종료 정리를 사용하면 다음을 얻는다.

$$0 = E_0 S^2_{N\wedge n} - E_0(N \wedge n)$$

단조 수렴 정리는 $E_0(N \wedge n) \uparrow E_0 N$임을 함의한다. 유계 수렴 정리 및 $x = 0$에 대한 (a)의 결과는 다음을 함의한다.

$$E_0 S^2_{N\wedge n} \to a^2 \frac{b}{b-a} + b^2 \frac{-a}{b-a} = -ab \left[\frac{-a}{b-a} + \frac{b}{b-a}\right] = -ab$$

이는 증명을 완료한다. □