Martingales and Almost Sure Convergence

김지수 (Jisu KIM)

확률론 2 (Probability Theory 2), 2025 2nd semester

Some parts of this lecture note are from the lecture notes from Prof. Alessandro Rianldo's "Advanced Probability Overview".

Martingales (마팅게일)

Martingales are elegant and powerful tools to study sequences of dependent random variables. It is originated from gambling, where a gambler can adjust the bet according to the previous results.

Definition. An increasing sequence of sub $\sigma-$fields $\{\mathcal{F}_n : n = 1, 2, \cdots\}$ of $\mathcal{F}$ is called a filtration. We call $\{X_n\}$ is adapted to $\{\mathcal{F}_n\}$ if $X_n \in \mathcal{F}_n$ for all n.

Definition. If $\{X_n\}$ is a sequence of random variables with

(i) $\mathbb{E}|X_n| < \infty$

(ii) $X_n \in \mathcal{F}_n$

(iii) $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$ a.s. for all n

then $\{X_n\}$ is a martingale (마팅게일) (with respect to $\mathcal{F}_n$)

Definition. If (iii) is replaced with $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$, it is a submartingale.
If (iii) is replaced with $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \leq X_n$, it is a supermartingale.

Remark. If $\{X_n, \mathcal{F}_n\}$ is a submartingale, $\{-X_n, \mathcal{F}_n\}$ is a supermartingale.
If $\{\mathcal{F}_n\}$ is not specified when a martingale $\{X_n\}$ is defined, we let $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$.
If $\{\mathcal{G}_n\}$ is a filtration such that $\mathcal{G}_n \subset \mathcal{F}_n$ and $X_n \in \mathcal{G}_n$, then $\{X_n, \mathcal{G}_n\}$ is also a martingale.

We begin by describing three examples related to random walk. Let $\xi_1, \xi_2, \cdots$ be independent and identically distributed, with $\mathbb{E}|\xi_i| < \infty$, and let $\mu = \mathbb{E}\xi_i$. Let S_0 be a constant, and $S_n = S_0 + \xi_1 + \cdots + \xi_n$. Let $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ for $n \geq 1$ and take $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Example ([1, Example 4.2.1])**.** Linear Martingale. If $\mu = 0$, then $\{S_n\}$ is a martingale with respect to $\mathcal{F}_n$.
To prove this, we observe that $S_n \in \mathcal{F}_n$, $\mathbb{E}|S_n| < \infty$, and ξ_{n+1} is independent of $\mathcal{F}_n$, so using the linearity of conditional expectation,

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

If $\mu \leq 0$, then $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) \leq X_n$, i.e., X_n is a supermartingale.
If $\mu \geq 0$ then S_n is a submartingale.
For any value of μ, by letting $\xi_i' = \xi_i - \mu$, we see that $S_n - n\mu$ is a martingale.

Example. Quadratic martiagale. Suppose now that $\mu = \mathbb{E}\xi_i = 0$ and $\sigma^2 = \mathrm{var}(\xi_i) < \infty$. In this case $S_n^2 - n\sigma^2$ is a martingale.

Since $(S_n + \xi_{n+1})^2 = S_n^2 + 2S_n\xi_{n+1} + \xi_{n+1}^2$ and ξ_{n+1} is independent of $\mathcal{F}_n$, we have

$$\mathbb{E}(S_{n+1}^2 - (n+1)\sigma^2 \mid \mathcal{F}_n) = S_n^2 + 2S_n\mathbb{E}(\xi_{n+1} \mid \mathcal{F}_n) + \mathbb{E}(\xi_{n+1}^2 \mid \mathcal{F}_n) - (n+1)\sigma^2 = S_n^2 - n\sigma^2.$$

Example. Exponential martingale. Let $Y_1, Y_2, \ldots$ be nonnegative i.i.d. random variables with $\mathbb{E}Y_m = 1$. If $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$ then

$$M_n = \prod_{m \le n} Y_m$$

defines a martingale. To prove this note that

$$\mathbb{E}(M_{n+1} \mid \mathcal{F}_n) = M_n\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) = M_n.$$

Suppose now that $Y_i = e^{\theta\xi_i}$ and $\varphi(\theta) = \mathbb{E}e^{\theta\xi_i} < \infty$. $Y_i = \exp(\theta\xi_i)/\varphi(\theta)$ has mean 1 so $\mathbb{E}Y_i = 1$ and

$$M_n = \prod_{i=1}^{n} Y_i = \frac{\exp(\theta S_n)}{\varphi(\theta)^n}$$

is a martingale.

Theorem ([1, Theorem 4.2.4, 4.2.5])**.** *(i) If X_n is a supermartingale then for $n > m$, $\mathbb{E}(X_n|\mathcal{F}_m) \le X_m$*

(ii) If X_n is a submartingale then for $n > m$, $\mathbb{E}(X_n|\mathcal{F}_m) \ge X_m$

(iii) If X_n is a martingale (마팅게일) then for $n > m$, $\mathbb{E}(X_n|\mathcal{F}_m) = X_m$

Proof. (i)

The definition gives the result for $n = m + 1$. Suppose $n = m + k$ with $k \ge 2$. By [1, Theorem 4.1.2],

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

by the definition of supermartingale and [1, Theorem 4.1.9 (b)]. The desired result now follows by induction.

(ii)

Note that $-X_n$ is a supermartingale and use [1, Theorem 4.1.9 (a)].

(iii)

Observe that X_n is both a supermartingale and a submartingale. □

Remark. The idea in the proof of (ii) and (iii) will be used many times below. To keep from repeating ourselves, we will just state the result for either supermartingales or submartingales and leave it to the reader to translate the result for the other two.

Theorem ([1, Theorem 4.2.6])**.** *Let X_n be a martingale (마팅게일) with respect to $\mathcal{F}_n$ and φ be a convex function (볼록함수) with $\mathbb{E}|\varphi(X_n)| < \infty$ for all n. Then, $\varphi(X_n)$ is a submartingale with respect to $\mathcal{F}_n$. Consequently, if $p \ge 1$ and $\mathbb{E}|X_n|^p < \infty$ for all n, then $|X_n|^p$ is a submartingale with respect to $\mathcal{F}_n$.*

Proof. By Jensen's inequality and the definition,

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

□

Theorem ([1, Theorem 4.2.7]). *Let X_n be a submartingale with respect to $\mathcal{F}_n$, and φ be an increasing convex function(볼록증가함수) with $\mathbb{E}|\varphi(X_n)| < \infty$ for all n. Then $\varphi(X_n)$ is a submartingale with respect to $\mathcal{F}_n$. Consequently,*

(i) If X_n is a submartingale then $(X_n - a)^+$ is a submartingale.

(ii) If X_n is a supermartingale then $X_n \wedge a$ is a supermartingale.

Proof. By Jensen's inequality and the assumptions,

$$\mathbb{E}(\varphi(X_{n+1}) \mid \mathcal{F}_n) \geq \varphi(\mathbb{E}(X_{n+1} \mid \mathcal{F}_n)) \geq \varphi(X_n).$$

$\square$

Definition. A sequence of random variables $\{H_n : \ n = 1, 2, \cdots\}$ is said to be predictable relative to the filtration $\{\mathcal{F}_n\}$ if $H_n \in \mathcal{F}_{n-1}$ for all $n \geq 1$.

In words, the value of H_n may be predicted (with certainty) from the information available at time $n-1$. In this section, we will be thinking of H_n as the amount of money a gambler will bet at time n. This can be based on the outcomes at times $1, \ldots, n-1$ but not on the outcome at time n.

Once we start thinking of H_n as a gambling system, it is natural to ask how much money we would make if we used it. Let X_n be the net amount of money you would have won at time n if you had bet one dollar each time. If you bet according to a gambling system H then your winnings at time n would be as follows:

Definition. Let X_n be a (sub, super) martingale with respect to $\mathcal{F}_n$ and let $\{H_n\}$ be predictable. Define

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

since if at time m you have wagered \$3 the change in your fortune would be 3 times that of a person who wagered \$1. Alternatively you can think of X_m as the value of a stock and H_m the number of shares you hold from time $m-1$ to time m.

Suppose now that $\xi_m = X_m - X_{m-1}$ have $P(\xi_m = 1) = p$ and $P(\xi_m = -1) = 1 - p$. A famous gambling system called the **martingale** is defined by $H_1 = 1$ and for $n \geq 2$,

$$H_n = \begin{cases} 2H_{n-1}, & \text{if } \xi_{n-1} = -1, \\ 1, & \text{if } \xi_{n-1} = 1. \end{cases}$$

In words, we double our bet when we lose, so that if we lose k times and then win, our net winnings will be 1. To see this consider the following concrete situation:

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

This system seems to provide us with a "sure thing" as long as $P(\xi_m = 1) > 0$. However, the next result says there is no system for beating an unfavorable game.

Theorem. *Let X_n be a (sub, super) martingale with respect to $\mathcal{F}_n$ and let $\{H_n\}$ be predictable. Suppose $(H \bullet X)_n$ is integrable (this holds in particular when each H_n is bounded).*

Then $\{(H \bullet X)_n, \ \mathcal{F}_n\}$ is a martingale.

If $H_n \geq 0$, it is a (sub, super) martingale.

Proof. Using the fact that conditional expectation is linear, $(H \cdot X)_n \in \mathcal{F}_n$, $H_n \in \mathcal{F}_{n-1}$, and [1, Theorem 4.1.14], we have

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

since $\mathbb{E}((X_{n+1} - X_n) \mid \mathcal{F}_n) \leq 0$ and $H_{n+1} \geq 0$. □

Remark. $(H \bullet X)_n$ is called an integral transformation of H_n with respect to X_n, and is a discrete-time version of the famous Ito's stochastic integral

$(H \bullet X)_t = \int_0^t H_s dX_s$

or the equivalent differential equation form

$d(H \bullet X)_t = H_t dX_t$ with $(H \bullet X)_0 = 0$

We will now consider a very special gambling system: bet \$1 at each time $n \leq N$ then stop playing.

Definition. A nonnegative integer valued random variable $N : \Omega \to \{0, 1, \dots\} \cup \{\infty\}$ is said to be a stopping time (정지시간) if $\{N = n\} \in \mathcal{F}_n$ for all $n < \infty$, i.e., the decision to stop at time n must be measurable with respect to the information known at that time.

Theorem ([1, Theorem 4.2.9])**.** *If N is a stopping time (정지시간) and X_n is a supermartingale, then $X_{N \wedge n}$ is a supermartingale.*

Proof. If we let $H_n = 1_{\{N \geq n\}}$, then $\{N \geq n\} = \{N \leq n-1\}^c \in \mathcal{F}_{n-1}$, so H_n is predictable. Then [1, Theorem 4.2.8] implies that

$$(H \cdot X)_n = X_{N \wedge n} - X_0$$

is a supermartingale. Since the constant sequence $Y_n = X_0$ is a supermartingale and the sum of two supermartingales is also supermartingale, we have that $X_{N \wedge n}$ is a supermartingale. □

Almost sure convergence (거의 확실한 수렴)

Submartingale is a stochastic version of monotone increasing sequence. Hence, we expect it converges a.s. if $\sup_n \mathbb{E} X_n^+ < \infty$.

Although [1, Theorem 4.2.8] implies that you cannot make money with gambling systems, you can prove theorems with them. Let $a < b$ and $\{X_n\}$ be stochastic process. Let $N_0 = 1$ and

$$\begin{aligned}N_1 &= \min\{m \geq 1 : \ X_m \leq a\} \\ N_2 &= \min\{m > N_1 : \ X_m \geq b\} \\ &\vdots \\ N_{2k-1} &= \min\{m > N_{2k-2} : X_m \leq a\} \\ N_{2k} &= \min\{m > N_{2k-1} : \ X_m \geq b\}.\end{aligned}$$

Then N_j are stopping times and

$$\{N_{2k-1} < m \leq N_{2k}\} = \{N_{2k-1} \leq m-1\} \cap \{N_{2k} \leq m-1\}^{\complement} \in \mathcal{F}_{m-1}.$$

and so

$$H_m = \begin{cases} 1, & \text{if } N_{2k-1} < m \le N_{2k} \text{ for some } k, \\ 0, & \text{otherwise,} \end{cases}$$

defines a predictable sequence. $X_{N_{2k-1}} \le a$ and $X_{N_{2k}} \ge b$, so between times N_{2k-1} and N_{2k}, X_m crosses from below a to above b. H_m is a gambling system that tries to take advantage of these "upcrossings." In stock market terms, we buy when $X_m \le a$ and sell when $X_m \ge b$, so every time an upcrossing is completed, we make a profit of $\ge (b-a)$.

Finally, let $U_n = \sup\{k : \ N_{2k} \le n\}$ be the number of upcrossings completed by time n.

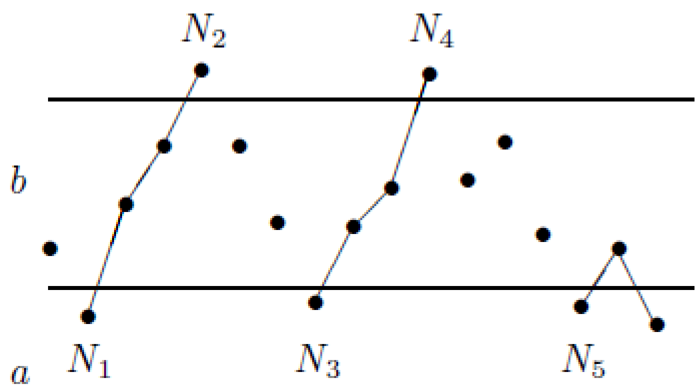

Figure 1: Upcrossings of (a, b). Lines indicate increments that are included in $(H \cdot X)_n$. In Y_n the points $< a$ are moved up to a.

Definition.

Theorem ([1, Theorem 4.2.10])**.** *Upcrossing inequality*

If $\{X_n\}$ is a submartingale, then

$$(b-a)\mathbb{E}(U_n) \le \mathbb{E}(X_n - a)^+ - \mathbb{E}(X_0 - a)^+.$$

(not covered in class). Let $Y_m = a + (X_m - a)^+$. By [1, Theorem 4.2.7], Y_m is a submartingale. Clearly, it upcrosses $[a, b]$ the same number of times that X_m does, and we have $(b-a)U_n \le (H \cdot Y)_n$, since each upcrossing results in a profit $\ge (b-a)$ and a final incomplete upcrossing (if there is one) makes a nonnegative contribution to the right-hand side. It is for this reason we had to replace X_m by Y_m.

Let $K_m = 1 - H_m$. Clearly, $Y_n - Y_0 = (H \cdot Y)_n + (K \cdot Y)_n$, and it follows from [1, Theorem 4.2.8] that $\mathbb{E}(K \cdot Y)_n \ge \mathbb{E}(K \cdot Y)_0 = 0$, so $\mathbb{E}(H \cdot Y)_n \le \mathbb{E}(Y_n - Y_0)$, proving the desired inequality. $\square$

The upcrossing lemma says that a submartingale cannot cross a fixed nondegenerate interval very often with high probability. If the submartingale were to cross an interval infinitely often, then its lim sup and lim inf would have to be different and it couldn't converge.

Theorem ([1, Theorem 4.2.11])**.** *Martingale convergence theorem (마팅게일 수렴정리).*

If $\{X_n\}$ is a submartingale with $\sup_n \mathbb{E}X_n^+ < \infty$, then as $n \to \infty$, X_n converges a.s. (거의 확실한 수렴) to a limit X with $\mathbb{E}|X| < \infty$.

Proof. Since $(X-a)^+ \le X^+ + |a|$, [1, Theorem 4.2.10] implies that

$$\mathbb{E}U_n \le (|a| + \mathbb{E}X_n^+)/(b-a).$$

As $n \uparrow \infty$, $U_n \uparrow U$, where U is the number of upcrossings of $[a, b]$ by the whole sequence. So if $\sup \mathbb{E}X_n^+ < \infty$ then $\mathbb{E}U < \infty$ and hence $U < \infty$ a.s.. Since $\liminf X_n < a < b < \limsup X_n$ implies that $U = \infty$, this implies that

$$P(\liminf X_n < a < b < \limsup X_n) = 0.$$

Since this holds for all rational a and b,

$$\bigcup_{a,b \in \mathbb{Q}} \{\liminf X_n < a < b < \limsup X_n\}$$

has probability 0 and hence $\limsup X_n = \liminf X_n$ a.s., i.e., $\lim X_n$ exists a.s.. Fatou's lemma guarantees $\mathbb{E}X^+ \leq \liminf \mathbb{E}X_n^+ < \infty$, so $X < \infty$ a.s. To see $X > -\infty$, we observe that

$$\mathbb{E}X_n^- = \mathbb{E}X_n^+ - \mathbb{E}X_n \leq \mathbb{E}X_n^+ - \mathbb{E}X_0$$

(since X_n is a submartingale), so another application of Fatou's lemma shows

$$\mathbb{E}X^- \leq \liminf_{n \to \infty} \mathbb{E}X_n^- \leq \sup_n \mathbb{E}X_n^+ - \mathbb{E}X_0 < \infty.$$

□

An important special case of this theorem is:

Theorem ([1, Theorem 4.2.12])**.** *If $X_n \geq 0$ is a supermartingale, then $X_n \to X$ a.s. (거의 확실한 수렴) and $\mathbb{E}X \leq \mathbb{E}X_0$.*

Proof. $Y_n = -X_n \leq 0$ is a submartingale with $\mathbb{E}Y_n^+ = 0$. Since $\mathbb{E}X_0 \geq \mathbb{E}X_n$, the inequality follows from Fatou's lemma. □

We first give two "counterexamples."

Example ([1, Example 4.2.13])**.** The first shows that the assumptions of [1, Theorem 4.2.12] (or [1, Theorem 4.2.11]) do not guarantee convergence in L^1 (L^1 수렴).

Let $S_0 = 1$ and $\{S_n,\ n \geq 1\}$ be i.i.d. symmetric simple random walk. That is, $S_n = S_{n-1} + \xi_n$ where $\xi_1, \xi_2, \cdots$ are i.i.d. with $P(\xi_i = 1) = P(\xi_i = -1) = \frac{1}{2}$.

In fact, S_n does not converge and $\limsup S_n = \infty$, by [1, Exercise 5.4.1].

Let $N = \min\{n :\ S_n = 0\}$ and let $X_n = S_{N \wedge n}$. Then [1, Theorem 4.2.9] implies that X_n is a nonnegative martingale.

Hence [1, Theorem 4.2.12] implies that X_n converges to a limit X_∞ a.s. and $\mathbb{E}|X_\infty| < \infty$. In fact, $X_\infty = 0$ a.s., since convergence to $k > 0$ is impossible. (If $X_n = k > 0$ then $X_{n+1} = k \pm 1$.)

Since $\mathbb{E}X_n = \mathbb{E}X_0 = 1$ for all n and $X_\infty = 0$, $X_n \nrightarrow X$ in L_1.

The above example is an important counterexample to keep in mind as you read the rest of this chapter. The next one is not as important.

Example ([1, Example 4.2.14])**.** We will now give an example of a martingale with $X_k \to 0$ in probability (확률수렴) but not a.s. (거의 확실한 수렴).

Let $X_0 = 0$. When $X_{k-1} = 0$, let

$$X_k = \begin{cases} 1 & \text{with probability } 1/2k, \\ -1 & \text{with probability } 1/2k, \\ 0 & \text{with probability } 1 - 1/k. \end{cases}$$

When $X_{k-1} \neq 0$, let

$$X_k = \begin{cases} kX_{k-1} & \text{with probability } 1/k, \\ 0 & \text{with probability } 1 - 1/k. \end{cases}$$

From the construction, $P(X_k = 0) = 1 - 1/k$ so $X_k \to 0$ in probability.

On the other hand, the second Borel–Cantelli lemma implies $P(X_k = 0 \text{ for all } k \geq K) = 0$, and values in $(-1, 1) \setminus \{0\}$ are impossible, so X_k does not converge to 0 a.s.

Examples

Bounded Increments

Our first result shows that martingales with bounded increments either converge or oscillate between $+\infty$ and $-\infty$.

Theorem ([1, Theorem 4.3.1])**.** *Let $X_1, X_2, \ldots$ be a martingale with $|X_{n+1} - X_n| \leq M < \infty$. Let*

$$C = \{\lim X_n \text{ exists and is finite}\}, \quad D = \{\limsup X_n = +\infty \text{ and } \liminf X_n = -\infty\}.$$

Then $P(C \cup D) = 1$.

Proof. Since $X_n - X_0$ is a martingale, we can without loss of generality suppose $X_0 = 0$. Let $0 < K < \infty$ and let $N = \inf\{n : X_n \leq -K\}$. $X_{n \wedge N}$ is a martingale with $X_{n \wedge N} \geq -K - M$ a.s., so applying [1, Theorem 4.2.12] to $X_{n \wedge N} + K + M$ shows $\lim X_n$ exists on $\{N = \infty\}$. Letting $K \to \infty$, we see that the limit exists on $\{\liminf X_n > -\infty\}$. Applying the last conclusion to $-X_n$, we see that $\lim X_n$ exists on $\{\limsup X_n < \infty\}$ and the proof is complete. □

Theorem ([1, Theorem 4.3.1])**.** *Doob's decomposition. Any submartingale X_n, $n \geq 0$, can be written in a unique way as*

$$X_n = M_n + A_n,$$

where M_n is a martingale and A_n is a predictable increasing sequence with $A_0 = 0$.

Proof. We want $X_n = M_n + A_n$, $\mathbb{E}(M_n \mid \mathcal{F}_{n-1}) = M_{n-1}$, and $A_n \in \mathcal{F}_{n-1}$. So we must have

$$\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(M_n \mid \mathcal{F}_{n-1}) + \mathbb{E}(A_n \mid \mathcal{F}_{n-1}) = M_{n-1} + A_n = X_{n-1} - A_{n-1} + A_n,$$

and it follows that

$$A_n - A_{n-1} = \mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - X_{n-1}.$$

Since $A_0 = 0$, we have

$$A_n = \sum_{m=1}^{n} \mathbb{E}(X_m - X_{m-1} \mid \mathcal{F}_{m-1}).$$

To check that our recipe works, we observe that $A_n - A_{n-1} \geq 0$ since X_n is a submartingale and $A_n \in \mathcal{F}_{n-1}$. To prove that $M_n = X_n - A_n$ is a martingale, we note that using $A_n \in \mathcal{F}_{n-1}$ and the relations above,

$$\mathbb{E}(M_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_n - A_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - A_n = X_{n-1} - A_{n-1} = M_{n-1},$$

which completes the proof. □

Theorem ([1, Theorem 4.3.4]). *Second Borel–Cantelli lemma, II. Let $\mathcal{F}_n$, $n \geq 0$ be a filtration with $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and let B_n, $n \geq 1$ be a sequence of events with $B_n \in \mathcal{F}_n$. Then*

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

Proof. If we let $X_0 = 0$ and $X_n = \sum_{m \leq n} 1_{B_m}$, then X_n is a submartingale. Doob's decomposition [1, Theorem 4.3.2\] implies

$$A_n = \sum_{m=1}^{n} E(1_{B_m} \mid \mathcal{F}_{m-1}),$$

so if $M_0 = 0$ and

$$M_n = \sum_{m=1}^{n} \Big(1_{B_m} - P(B_m \mid \mathcal{F}_{m-1})\Big), \qquad n \geq 1,$$

then M_n is a martingale with $|M_n - M_{n-1}| \leq 1$.

Using the notation of [1, Theorem 4.3.1] we have:

On C,

$$\sum_{n=1}^{\infty} 1_{B_n} = \infty \quad \text{if and only if} \quad \sum_{n=1}^{\infty} P(B_n \mid \mathcal{F}_{n-1}) = \infty.$$

On D,

$$\sum_{n=1}^{\infty} 1_{B_n} = \infty \quad \text{and} \quad \sum_{n=1}^{\infty} P(B_n \mid \mathcal{F}_{n-1}) = \infty.$$

Since $P(C \cup D) = 1$, the result follows. □

Polya's Urn Scheme

An urn contains r red and g green balls. At each time we draw a ball out, then replace it, and add c more balls of the color drawn. Let X_n be the fraction of green balls after the nth draw.

We check that X_n is a martingale. Suppose there are i red balls and j green balls at time n, then

$$X_{n+1} = \begin{cases} \dfrac{j+c}{i+j+c} & \text{with probability } \dfrac{j}{i+j}, \\ \dfrac{j}{i+j+c} & \text{with probability } \dfrac{i}{i+j}, \end{cases}$$

so we have

$$E\left[X_{n+1} | X_n = \frac{j}{i+j}\right] = \frac{j+c}{i+j+c} \cdot \frac{j}{i+j} + \frac{j}{i+j+c} \cdot \frac{i}{i+j} = \frac{(j+c+i)j}{(i+j+c)(i+j)} = \frac{j}{i+j}.$$

Since $X_n \geq 0$, Martingale convergence theorem (마팅게일 수렴정리) [1, Theorem 4.2.12] implies that $X_n \to X_\infty$ a.s..

To compute the distribution of the limit X_∞, we observe (a) the probability of getting green on the first m

draws then red on the next $\ell = n - m$ draws is

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

and (b) any other outcome of the first n draws with m green draws and ℓ red draws has the same probability since the denominator remains the same and the numerator is permuted. Hence, by letting G_n be the number of green balls after the nth draw has been completed and the new ball has been added, then it follows from (a) and (b) that

$$P(G_n = mc+g) = \binom{n}{m} \frac{g}{g+r} \cdot \frac{g+c}{g+r+c} \cdot \cdots \cdot \frac{g+(m-1)c}{g+r+(m-1)c} \cdot \frac{r}{g+r+mc} \cdot \frac{r+c}{g+r+(m+1)c} \cdot \cdots \cdot \frac{r+(\ell-1)c}{g+r+(n-1)c}.$$

Consider the special case $c = 1$, $g = 1$, and $r = 1$, then

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

so

$$X_\infty \sim \mathrm{Unif}(0,1).$$

If we suppose that $c = 1$, $g = 2$, and $r = 1$, then

$$P(G_n = m+2) = \frac{n!}{m!(n-m)!} \frac{(m+1)!(n-m)!}{(n+2)!/2} \to 2x \quad \text{if } n \to \infty \text{ and } m/n \to x,$$

so

$$X_\infty \sim \mathrm{Beta}(2,1).$$

In general, the distribution of X_∞ has density

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

so

$$X_\infty \sim \mathrm{Beta}\left(\frac{g}{c}, \frac{r}{c}\right).$$

In [1, Example 4.5.6], the limit behavior changes drastically if, in addition to the c balls of the color chosen, we always add one ball of the opposite color.

Branching Processes

Let ξ_i^n, $i, n \geq 1$, be i.i.d. nonnegative integer-valued random variables. Define a stochastic process $\{Z_n\}_{n\geq 0}$ by $Z_0 = 1$ and

$$Z_{n+1} = \begin{cases} \xi_1^{n+1} + \cdots + \xi_{Z_n}^{n+1}, & \text{if } Z_n > 0, \\ 0, & \text{if } Z_n = 0. \end{cases}$$

This process is called a *Galton–Watson process*. The idea behind the definitions is that Z_n is the number of individuals in the nth generation, and each member of the nth generation gives birth independently to an identically distributed number of children. $p_k = P(\xi_i^n = k)$ is called the offspring distribution.

One of the fundamental question for $\{Z_n\}$ is to calculate $P(Z_n = 0 \text{ for some } n)$ (i.e. extinction probability).

Lemma. *[1, Lemma 4.3.9\] Let $\mathcal{F}_n = \sigma(\xi_i^m : i \geq 1, 1 \leq m \leq n)$ and $\mu = \mathbb{E}\xi_i^m \in (0,\infty)$. Then Z_n/μ^n is a martingale w.r.t. $\mathcal{F}_n$.*

Proof. Clearly, $Z_n \in \mathcal{F}_n$. Using [1, Theorem 4.1.2], we conclude that on $\{Z_n = k\}$,

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

where in the second equality we used the fact that the ξ_j^{n+1} are independent of $\mathcal{F}_n$. $\square$

Z_n/μ^n is a nonnegative martingale, so Martingale convergence theorem (마팅게일 수렴정리) [1, Theorem 4.2.12] implies $Z_n/\mu^n \to$ a limit a.s. We begin by identifying cases when the limit is trivial.

Theorem. *[1, Theorem 4.3.10] If $\mu < 1$ then $Z_n = 0$ for all n sufficiently large, so $Z_n/\mu^n \to 0$.*

Proof. $\mathbb{E}(Z_n/\mu^n) = \mathbb{E}(Z_0) = 1$, so $\mathbb{E}(Z_n) = \mu^n$. Now $Z_n \geq 1$ on $\{Z_n > 0\}$, so

$$P(Z_n > 0) \leq \mathbb{E}(Z_n; Z_n > 0) = \mathbb{E}(Z_n) = \mu^n \to 0$$

exponentially fast if $\mu < 1$. $\square$

The last answer should be intuitive. If each individual on the average gives birth to less than one child, the species will die out. The next result shows that after we exclude the trivial case in which each individual has exactly one child, the same result holds when $\mu = 1$.

Theorem. *[1, Theorem 4.3.11] If $\mu = 1$ and $P(\xi_i^m = 1) < 1$ then $Z_n = 0$ for all n sufficiently large.*

Proof. When $\mu = 1$, Z_n is itself a nonnegative martingale. Since Z_n is integer valued and by Martingale convergence theorem (마팅게일 수렴정리) [1, Theorem 4.2.12], Z_n converges to an a.s. finite limit Z_∞, we must have $Z_n = Z_\infty$ for large n. If $P(\xi_i^m = 1) < 1$ and $k > 0$ then $P(Z_n = k \text{ for all } n \geq N) = 0$ for any N, so we must have $Z_\infty \equiv 0$. $\square$

When $\mu \leq 1$, the limit of Z_n/μ^n is 0 because the branching process dies out. Our next step is to show that if $\mu > 1$ then $P(Z_n > 0 \text{ for all } n) > 0$.

Definition. For a nonnegative integer valued random variable X, i.e., $X \in \{0, 1, \ldots\}$, then the (probability) generating function of X is a function $\varphi : I \subset \mathbb{R} \to \mathbb{R}$ such that

$$\varphi(s) = \mathbb{E}\left[s^X\right] = \sum_{k=0}^{\infty} P(X = k)s^k.$$

We consider the generating function $\varphi : [0, 1] \to \mathbb{R}$ for ξ_i^m, which is the generating function for the offspring distribution $p_k = P(\xi_i^m = k)$.

Theorem. *[1, Theorem 4.3.12] Suppose $\mu > 1$. If $Z_0 = 1$ then $P(Z_n = 0 \text{ for some } n) = \rho$, the only solution of $\varphi(\rho) = \rho$ in $[0, 1)$.*

Proof. $\varphi(1) = 1$. Differentiating and referring to [1, Theorem A.5.3] for the justification gives for $s < 1$,

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

so φ is increasing. We may have $\varphi(s) = \infty$ when $s > 1$ so we have to work carefully.

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

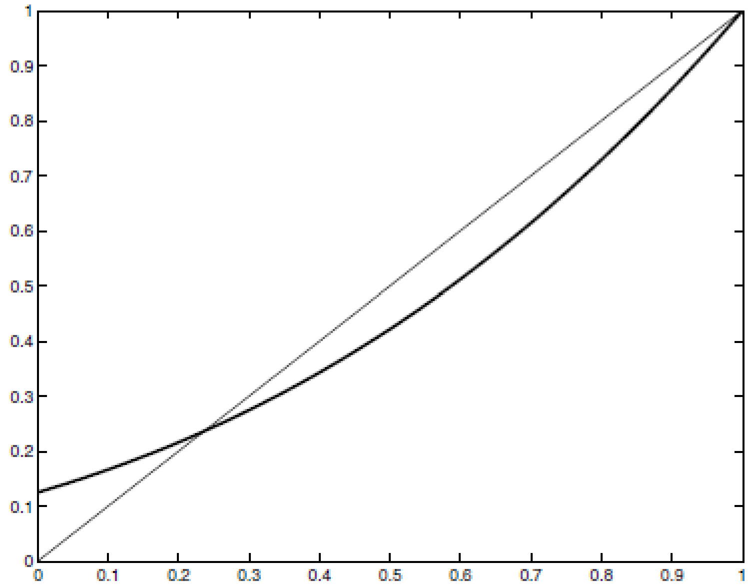

Figure 2: Generating function for Binomial$(3, 1/2)$.

Integrating we have

$$\varphi(1) - \varphi(1-h) = \int_{1-h}^{1} \varphi'(s)\, ds \sim \mu h \quad \text{as } h \to 0,$$

so if h is small $\varphi(1-h) < 1-h$. $\varphi(0) \geq 0$ so there must be a solution of $\varphi(x) = x$ in $[0,1)$.

To prove uniqueness we note that for $s < 1$,

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

since $\mu > 1$ implies that $p_k > 0$ for some $k \geq 2$. Let ρ be the smallest solution of $\varphi(\rho) = \rho$ in $[0,1)$. Since $\varphi(1) = 1$ and φ is strictly convex we have $\varphi(x) < x$ for $x \in (\rho, 1)$ so there is only one solution of $\varphi(\rho) = \rho$ in $[0,1)$.

Combining the next two results will complete the proof.

(a) If $\theta_m = P(Z_m = 0)$ then

$$\theta_m = \sum_{k=0}^{\infty} p_k (\theta_{m-1})^k = \varphi(\theta_{m-1}).$$

(b) As $m \uparrow \infty$, $\theta_m \uparrow \rho$.

Proof of (a). If $Z_1 = k$, an event with probability p_k, then $Z_m = 0$ iff all k families die out in the remaining $m-1$ units of time, an independent event with probability θ_{m-1}^k. Summing over the disjoint possibilities for each k gives the desired result.

Proof of (b). Clearly $\theta_m = P(Z_m = 0)$ is increasing. To show by induction that $\theta_m \leq \rho$, note that $\theta_0 = 0 \leq \rho$, and if the result is true for $m-1$,

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

Taking limits in $\theta_m = \varphi(\theta_{m-1})$, we see $\theta_\infty = \varphi(\theta_\infty)$. Since $\theta_\infty \leq \rho$, it follows that $\theta_\infty = \rho$.

□

The last result shows that when $\mu > 1$, the limit of Z_n/μ^n has a chance of being nonzero. The best result on

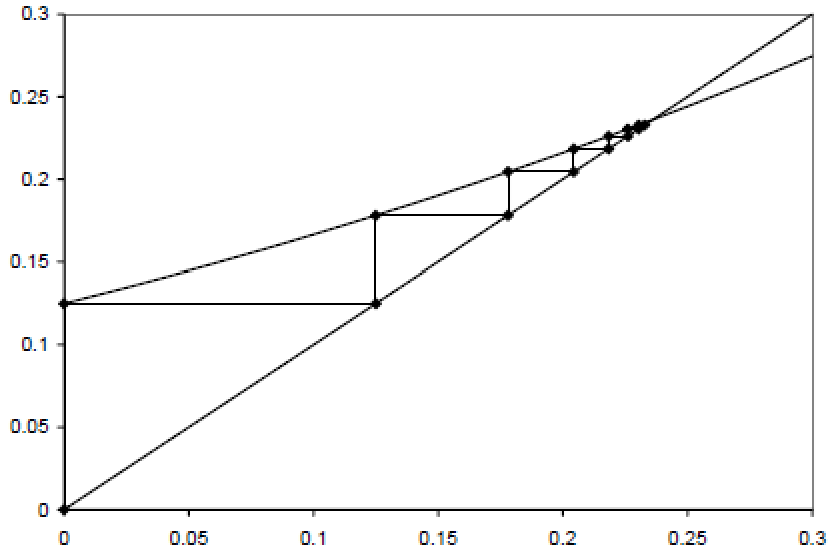

Figure 3: Iteration as in the proof for Binomial$(3, 1/2)$ generating function.

this question is due to Kesten and Stigum:

Theorem. *[1, Theorem 4.3.12] $W = \lim Z_n/\mu^n$ is not $\equiv 0$ if and only if*

$$\sum k p_k \log k < \infty.$$

For a proof, see Athreya and Ney (1972), p. 24–29. In the next lecture note, we will see that $\sum k^2 p_k < \infty$ (i.e., $var(\xi_i^m) < \infty$) is sufficient for a nontrivial limit.

Radon–Nikodym Derivatives

Let $(\Omega, \mathcal{F})$ be a measurable space, and let μ be a finite measure and ν a probability measure on $(\Omega, \mathcal{F})$. Let $\{\mathcal{F}_n\}$ be σ-fields with $\mathcal{F}_n \uparrow \mathcal{F}$ (i.e., $\sigma(\cup \mathcal{F}_n) = \mathcal{F}$). Let μ_n and ν_n be the restrictions of μ and ν to $\mathcal{F}_n$.

Definition. Let μ, ν be measure. We say ν is said to be absolutely continuous with respect to μ (abbreviated $\nu \ll \mu$) if $\mu(A) = 0$ implies $\nu(A) = 0$. We say μ and ν are mutually singular (abbreviated $\mu \perp \nu$) if there is a set A such that $\mu(A) = 0$ and $\nu(A^{\complement}) = 0$. We also say μ is singular with respect to ν.

Theorem. *[1, Theorem 4.3.5] Suppose $\mu_n \ll \nu_n$ for all n. Let $X_n = d\mu_n/d\nu_n$ and let $X = \limsup X_n$. Then*

$$\mu = \mu_r + \mu_s,$$

where

$$\mu_r(A) = \int_A X d\nu \qquad \textit{and} \qquad \mu_s(A) = \mu(A \cap \{X = \infty\}).$$

Remark. μ_r is a measure $\ll \nu$. Since Martingale convergence theorem (마팅게일 수렴정리) [1, Theorem 4.2.12] implies $\nu(X = \infty) = 0$, μ_s is singular with respect to ν. Thus $\mu = \mu_r + \mu_s$ gives the Lebesgue decomposition of μ (see [1, Theorem A.4.7]), and $X = d\mu_r/d\nu$, $[\nu]$-a.s.

Lemma. *[1, Lemma 4.3.6] X_n (defined on $(\Omega, \mathcal{F}, \nu)$) is a martingale with respect to $\mathcal{F}_n$.*

Proof. We observe that, by definition, $X_n \in \mathcal{F}_n$. Let $A \in \mathcal{F}_n$. Since $X_n \in \mathcal{F}_n$ and ν_n is the restriction of ν to $\mathcal{F}_n$,

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

Using the definition of X_n and [1, Exercise A.4.7],

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

the last equality holding since $A \in \mathcal{F}_n$ and μ_n is the restriction of μ to $\mathcal{F}_n$.

If $A \in \mathcal{F}_{m-1} \subset \mathcal{F}_m$, using the last result for $n = m$ and $n = m-1$ gives

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

so $E(X_m \mid \mathcal{F}_{m-1}) = X_{m-1}$.

Since X_n is a nonnegative martingale, Martingale convergence theorem (마팅게일 수렴정리) [1, Theorem 4.2.12] implies that $X_n \to X$ ν-a.s.

To see whether $\mu(A) = \int_A X d\nu + \mu(A \cap \{X = \infty\})$ holds, see [1, Section 4.3]. □

Kakutani dichotomy for infinite product measures

Let μ and ν be measures on sequence space $(\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}})$ that make the coordinates $\xi_n(\omega) = \omega_n$ independent. Let $F_n(x) = \mu(\xi_n \le x)$, $G_n(x) = \nu(\xi_n \le x)$ be the distribution functions of ξ_n under μ and ν, respectively.. Suppose $F_n \ll G_n$ and let $q_n = dF_n/dG_n$. To avoid a problem we will suppose $q_n > 0$, G_n-a.s.

Let $\mathcal{F}_n = \sigma(\xi_m : m \le n)$, let μ_n and ν_n be the restrictions of μ and ν to $\mathcal{F}_n$, and let

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

[1, Theorem 4.3.5] implies that $X_n \to X$ ν-a.s. Thanks to our assumption $q_n > 0$, G_n-a.s.. Now, note that

$$\sum_{m=1}^{\infty} \log(q_m) > -\infty$$

is a tail event, so the Kolmogorov 0–1 law implies

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

and it follows from [1, Theorem 4.3.5] that either $\mu \ll \nu$ or $\mu \perp \nu$. The next result gives a concrete criterion for which of the two alternatives occurs.

Theorem. *[1, Theorem 4.3.5]*

$$\begin{aligned} \mu \ll \nu \qquad & \textit{if} \prod_{m=1}^{\infty} \int \sqrt{q_m} dG_m > 0, \\ \mu \perp \nu \qquad & \textit{if} \prod_{m=1}^{\infty} \int \sqrt{q_m} dG_m = 0. \end{aligned}$$

For the proof, see [1, Section 4.3].

References

[1] Rick Durrett. *Probability—theory and examples*, volume 49 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2019. Fifth edition.