

# Chapter 1

## 1.1 Martingales

p.s. for a probability space.

r.v. for a random variable.

### Definition 1.1

For a p.s.  $(\Omega, \mathcal{F}_0, P)$  a  $\sigma$ -field  $\mathcal{F} \subset \mathcal{F}_0$  and a r.v.  $X \in \mathcal{F}_0$  with  $E|X| < \infty$ . We define the conditional expectation of  $X$  given  $\mathcal{F}$ ,  $E(X|\mathcal{F})$  to be any r.v.  $Y$  that has

a.  $Y \in \mathcal{F}$ .

b.  $\int_A X dP = \int_A Y dP$  for all  $A \in \mathcal{F}$ . and  $Y$  is said to be a version of  $E(X|\mathcal{F})$ .



### Lemma 1.1

If  $Y$  satisfies (a),(b) above, then it is integrable.



### Proof

We know

$$\int_{\{Y>0\}} Y dP = \int_{\{Y>0\}} X dP < \infty \quad \int_{\{Y<0\}} Y dP = \int_{\{Y<0\}} X dP < \infty$$

and hence  $\int |Y| dP$  finite.

### Lemma 1.2

If  $Y'$  also satisfies (a),(b) in Def.1.1., then  $Y = Y'$  a.s.



### Proof

Assume  $E_n = \{Y' - Y > n^{-1}\}$ ,  $F_n = \{Y - Y' > n^{-1}\}$ ,  $n \in \mathbb{N}$ , then we know

$$n^{-1}P(E_n) \leq \int_{E_n} (Y - Y') dP = \int_{E_n} Y dP - \int_{E_n} Y' dP = 0$$

and hence  $P(E_n) = 0$  for any  $n \in \mathbb{N}$ , similarly, we know  $P(F_n) = 0$  for any  $n \in \mathbb{N}$ , therefore,  $Y = Y'$  a.s.

### Theorem 1.1

If  $X_1 = X_2$  on  $B \in \mathcal{F}$  then  $E(X_1|\mathcal{F}) = E(X_2|\mathcal{F})$  a.s. on  $B$ .



### Proof

For any  $E \subset B$ , we have

$$0 = \int_{\{E(X_1|\mathcal{F}) - E(X_2|\mathcal{F}) > n^{-1}\} \cap E} (X_1 - X_2) dP \geq n^{-1}P(\{E(X_1|\mathcal{F}) - E(X_2|\mathcal{F}) > n^{-1}\} \cap E)$$

and the rest is similar.

### Theorem 1.2

$E(X|\mathcal{F})$  exists.



### Proof

Define  $\nu(E) = \int_E X dP$  for  $E \in \mathcal{F}$  and we know  $\nu \ll P$  and hence there exists  $Y \in \mathcal{F}$  such that

$$\int_E Y dP = \nu(E) = \int_E X dP$$

for all  $E \in \mathcal{F}$  by Radon-Nikodym's Theorem.

**Example 1.1** a. If  $X \in \mathcal{F}$ , then  $E(X|\mathcal{F}) = X$ .

b. If  $X$  is independent to  $\mathcal{F}$ , i.e.  $P(\{X \in B\} \cap A) = P(X \in B)P(A)$ , then  $X$  is independent to  $\chi_A$  for any  $A \in \mathcal{F}$  and hence  $E(X|\mathcal{F}) = EX$ .

c. Suppose  $\Omega_1, \Omega_2, \dots$  is a finite or infinite partition of  $\Omega$  into disjoint sets, with  $P(\Omega_i) > 0, i \geq 1$  and then let  $\mathcal{F} = \sigma(\Omega_1, \Omega_2, \dots)$  and then

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

d. Suppose  $X, Y$  have joint density  $f(x, y)$  i.e.,

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

then if  $E|g(X)| < \infty$ , then  $E(g(X)|Y) = h(Y)$ , where

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

on  $\{(x, y), \int f(x, y)dx > 0\}$ , and hence a.s.

e. Suppose  $X$  and  $Y$  are independent, let  $\phi$  be a function with  $E|\phi(X, Y)| < \infty$  and let  $g(x) = E(\phi(x, Y))$ , then  $E(\phi(X, Y)|X) = g(X)$ .

**Proof**

c. By the  $\pi - \lambda$  theorem, it suffices to show that

$$\int_A X dP = \int_A Y dP$$

for any  $A \in \{\bigcup_{1 \leq i \leq n} \Omega_i\}$  where  $Y$  was defined as above.

d. Firstly, we recall any simple function  $\phi \geq 0$  will cause  $\int \phi(x, y)dy$  is measurable since  $\int \phi(x, y)dy = \nu(E_y)$  when  $\phi = \chi_E$  and then we know for any  $g \geq 0$ ,  $\int g(x)f(x, y)dy$  is measurable and then  $\int g(x)f(x, y)dy$  is measurable for general  $g$ , then we will know  $h(Y) \in \sigma(Y)$ .

Consider  $A \in \sigma(Y)$ , where  $A = \{Y \in B\}$ , then

$$E(h(Y); A) = \int_{Y \in B} h(y)f(x, y)dx dy = \int_B \int h(y)f(x, y)dx dy = \int_B \int g(x)f(x, y)dx dy = E(g(X); A)$$

and the conclusion goes.

e. We know  $g(X) \in \sigma(X)$  and then for any  $A = \{X \in B\}$ , we will know

$$E(g(X); A) = \int_B g(x)dx = \int_B \int \phi(x, y)dy dx = E(\phi(X, Y); A)$$

and hence  $E(\phi(X, Y)|X) = g(X)$ .

### Definition 1.2

Denote

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

$$P(A|B) = P(A \cap B)/P(B)$$

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



### Theorem 1.3

For the first two parts, we assume  $E|X|, E|Y| < \infty$ .

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

(b) If  $X \leq Y$  then  $E(X|\mathcal{F}) \leq E(Y|\mathcal{F})$ .

(c) If  $X_n \geq 0$  and  $X_n \uparrow X$  with  $EX < \infty$  then  $E(X_n|\mathcal{F}) \uparrow E(X|\mathcal{F})$ .



### Theorem 1.4

If  $\phi$  is convex and  $E|X|, E|\phi(X)| < \infty$  then

$$\phi(E(X|\mathcal{F})) \leq E(\phi(X)|\mathcal{F})$$



**Proof**

Let  $S = \{(a, b) : a, b \in \mathbb{Q}, ax + b \leq \phi(x) \text{ for all } x\}$ , then  $\phi(x) = \sup\{ax + b : (a, b) \in S\}$ . And we know

$$E(\phi(X)|\mathcal{F}) \geq aE(X|\mathcal{F}) + b$$

and hence  $E(\phi(X)|\mathcal{F}) \geq \phi(E(X|\mathcal{F}))$ .

### Theorem 1.5

Conditional expectation is a contraction in  $L^p$ ,  $p \geq 1$ .



### Proof

By Theorem 1.5., we have  $|E|(X|\mathcal{F})|^p \leq E(|X|^p|\mathcal{F})$ , then we know

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

### Theorem 1.6

If  $\mathcal{F} \subset \mathcal{G}$  and  $E(X|\mathcal{G}) \in \mathcal{F}$ , then  $E(X|\mathcal{F}) = E(X|\mathcal{G})$ .



### Theorem 1.7

If  $\mathcal{F}_1 \subset \mathcal{F}_2$  then

$$(i) E(E(X|\mathcal{F}_1)|\mathcal{F}_2) = E(X|\mathcal{F}_1)$$

$$(ii) E(E(X|\mathcal{F}_2)|\mathcal{F}_1) = E(X|\mathcal{F}_1)$$



### Proof

For  $A \in \mathcal{F}_1$ , we know

$$\begin{aligned} \int_A E(E(X|\mathcal{F}_1)|\mathcal{F}_2) dP &= \int_A E(X|\mathcal{F}_1) dP = \int_A X dP \\ \int_A E(E(X|\mathcal{F}_2)|\mathcal{F}_1) dP &= \int_A E(X|\mathcal{F}_2) dP = \int_A X dP \end{aligned}$$

therefore, the equalities go.

### Theorem 1.8

If  $X \in \mathcal{F}$  and  $E|Y|, E|XY| < \infty$  then

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



### Proof

For any  $X, Y \geq 0$ , assume  $\phi_n \uparrow X$  simple, then we know  $\phi_n Y \uparrow XY$  and then

$$\int_A E(\chi_B Y|\mathcal{F}) = \int_A \chi_B Y dP = \int_{AB} Y dP = \int_{AB} E(Y|\mathcal{F}) dP = \int_A \chi_B E(Y|\mathcal{F})$$

for any  $A, B \in \mathcal{F}$  and hence  $E(\chi_B Y|\mathcal{F}) = \chi_B E(Y|\mathcal{F})$  for any  $B \in \mathcal{F}$ , therefore, we know  $E(\phi_n Y|\mathcal{F}) = \phi_n E(Y|\mathcal{F})$ .

By theorem 1.3 we know  $E(\phi_n Y|\mathcal{F}) \uparrow E(XY|\mathcal{F})$  and hence  $E(XY|\mathcal{F}) = XE(Y|\mathcal{F})$ , so for any  $X \in \mathcal{F}, E|Y| < \infty, E|XY| < \infty$ , we can consider the positive and negative parts and the conclusion goes.

### Theorem 1.9

Suppose  $EX^2 < \infty$ .  $E(X|\mathcal{F})$  is the variable  $Y \in \mathcal{F}$  that minimizes the "mean square error"  $E(X - Y)^2$ .



### Proof

If  $Z \in L^2(\mathcal{F})$ , then

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

then we know

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

and hence  $E(Z(X - E(X|\mathcal{F}))) = 0$  for any  $Z \in L^2(\mathcal{F})$ .

If  $Z = E(X|\mathcal{F}) - Y$ , then

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

and hence  $E(X - Y)^2$  are minimal when  $Y = E(X|\mathcal{F})$ .

### Definition 1.3

Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $X : (\Omega, \mathcal{F}) \rightarrow (S, \mathcal{S})$  and  $\mathcal{G}$  a  $\sigma$ -algebra contained by  $\mathcal{F}$ .  $\mu : \Omega \times \mathcal{S} \rightarrow [0, 1]$  is said to be a regular conditional distribution for  $X$  given  $\mathcal{G}$  if

- For each  $A$ ,  $\omega \rightarrow \mu(\omega, A)$  is a version of  $P(X \in A|\mathcal{G})$ .
- For a.e.  $\omega$ ,  $A \rightarrow \mu(\omega, A)$  is a probability measure on  $(S, \mathcal{S})$ .

When  $S = \Omega$  and  $X$  is the identity map,  $\mu$  is called a regular condition probability.



### Proposition 1.1

Suppose  $X$  and  $Y$  have a joint density  $f(x, y) > 0$ . If

$$\mu(y, A) = \int_A f(x, y)dx / \int f(x, y)dx$$

then  $\mu(Y(\omega), A)$  is a r.c.d for  $X$  given  $\sigma(Y)$ .



### Proof

Here we know  $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{R})$ , so we should check:

- $\mu(Y(\omega), A) = \int_A f(x, Y(\omega))dx / \int f(x, Y(\omega))dx$  is a version of  $P(X \in A|Y)$ .
- For a.e.  $\omega$ ,  $\mu_{Y(\omega)}(A) = \mu(Y(\omega), A)$  is a probability measure on  $(\mathbb{R}, \mathcal{R})$ .

To see the first claim, consider

$$\begin{aligned} \int_{Y \in B} P(X \in A|Y) dP &= \int_{Y \in B} \chi_{X \in A} dP = \int_B \int_A f(x, y) dx dy \\ &= \int_A \int_B f(x, y) dy dx \\ &= \int_B \int_A f(x, y) dx dy \\ &= \int_B \int_A \int f(x, y) dx / \int f(x, y) dx f(x, y) dx dy = \int_{Y \in B} \mu(Y(\omega), A) dP \end{aligned}$$

and the second claim is trivial.

### Theorem 1.10

Let  $\mu(\omega, A)$  be a r.c.d for  $X$  given  $\mathcal{F}$ . If  $f : (S, \mathcal{S}) \rightarrow (\mathbb{R}, \mathcal{R})$  has  $E|f(X)| < \infty$  then

$$E(f(x)|\mathcal{F}) = \int \mu(\omega, dx) f(x) \quad \text{a.s.}$$



### Proof

Consider  $f = \chi_A$  for some  $A$  mrb in  $\mathcal{R}$ , then  $\int \mu(\omega, dx) f(x) = \mu(\omega, A) = P(X \in A|\mathcal{F})$  and hence the equality holds for all simple functions, then the problem goes.

Here we skip some properties of regular conditional distribution and continue to martingale.

### Definition 1.4

$\mathcal{F}_n$  is a filtration, i.e. an increasing sequence of  $\sigma$ -fields. A sequence  $X_n$  is said to be adapted to  $\mathcal{F}_n$  if  $X_n \in \mathcal{F}_n$  for all  $n$ . If  $X_n$  is sequence with

- $E|X_n| < \infty$ .
- $X_n$  is adapted to  $\mathcal{F}_n$ .
- $E(X_{n+1}|\mathcal{F}_n) = X_n$  for all  $n$  then  $X$  is said to be a martingale (resp to  $\mathcal{F}_n$ ). If we replace the equality into  $\leq$  or  $\geq$ , then  $X$  is said to be a supermartingale or submartingale.



**Example 1.2** (Random walk) Let  $\xi_1, \xi_2, \dots$  be independent and i.i.d.,  $S_n = S_0 + \sum_{i=1}^n \xi_i$  where  $S_0$  is a constant.  $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$  and take  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ .

- If  $\mu = E\xi_i = 0$  then  $S_n, n \geq 0$  is a martingale with respect to  $\mathcal{F}_n$ .
- $\mu = E\xi_i = 0$  and  $\sigma^2 = \text{var}(\xi_i) < \infty$ , then  $S_n^2 - n\sigma^2$  is a martingale.

**Proof**

- Notice  $E|S_n| < \infty, n \geq 0$ , for any  $A \in \mathcal{F}_n$ , then notice

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

- Notice that  $E|S_n - n\sigma^2| < \infty$ , and

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

**Example 1.3** Let  $Y_1, Y_2, \dots$  be nonnegative i.i.d r.v.s with  $EY_m = 1$ . If  $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$ , then  $M_n = \prod_{m \leq n} Y_m$  defines a martingale.

Then assume  $\phi(\theta) = Ee^{\theta\xi_i}, Y_i = e^{\theta\xi_i}/\phi(\theta)$ , then we know  $M_n = e^{\theta S_n}/\phi(\theta)^n$ .

**Theorem 1.11**

If  $X_n$  is a (super-/sub-)martingale then for  $n > m$ ,  $E(X_n|\mathcal{F}_m) \leq (\geq / =) X_m$ .



**Proof** Notice

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

the rest proof is similar.

**Theorem 1.12**

If  $X_n$  is a martingale w.r.t.  $\mathcal{F}_n$  and  $\phi$  is a convex function with  $E|\phi(X_n)| < \infty$  for all  $n$  then  $\phi(X_n)$  is a submartingale w.r.t.  $\mathcal{F}_n$ . Consequently, if  $p \geq 1$  and  $E|X_n|^p < \infty$  for all  $n$ , then  $|X_n|^p$  is a submartingale w.r.t.  $\mathcal{F}_n$ .



**Proof** Notice

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

and the problem goes.

**Theorem 1.13**

If  $X_n$  is a submartingale w.r.t.  $\mathcal{F}_n$  and  $\phi$  is an increasing convex function with  $E|\phi(X_n)| < \infty$  for all  $n$ , then  $\phi(X_n)$  is a submartingale w.r.t.  $\mathcal{F}_n$ . Consequently

- If  $X_n$  is a submartingale then  $(X_n - a)^+$  is a submartingale.
- If  $X_n$  is a supermartingale then  $\min(X_n, a)$  is a supermartingale.



**Proof** Notice

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

then (a) is easy to be checked and hence (b) is correct.

**Definition 1.5**

Let  $\mathcal{F}_n, n \geq 0$  be a filtration.  $H_n, n \geq 1$  is said to be a predictable sequence if  $H_n \in \mathcal{F}_{n-1}$  for all  $n \geq 1$ .



**Definition 1.6**

We denote

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



**Theorem 1.14**

Let  $X_n, n \geq 0$  be a supermartingale. If  $H_n \geq 0$  is predictable and each  $H_n$  is bounded then  $(H \cdot X)_n$  is a supermartingale.



**Proof** Consider

$$E((H \cdot X)_{n+1} | \mathcal{F}_n) = E\left(\sum_{m=1}^{n+1} H_m (X_m - X_{m-1}) | \mathcal{F}_n\right) = (H \cdot X)_n + E(X_{n+1} | \mathcal{F}_n) - X_n \leq (H \cdot X)_n$$

**Definition 1.7**

A r.v.  $N$  is said to be a stopping time if  $\{N = n\}$  in  $\mathcal{F}_n$  for all  $n > \infty$ .

**Theorem 1.15**

If  $N$  is a stopping time and  $X_n$  is a supermartingale, then  $X_{N \wedge n}$  is a supermartingale.



**Proof** Consider

$$E(X_{N \wedge n+1} | \mathcal{F}_n) = E(X_{n+1} \chi_{N \geq n+1} + \sum_{k=0}^n X_k \chi_{N=k} | \mathcal{F}_n) \leq \chi_{N \geq n+1} X_n + \sum_{k=0}^n X_k \chi_{N=k} = X_{N \wedge n}$$

**Definition 1.8**

Suppose  $X_n, n \geq 0$  is a submartingale. Let  $a < b, N_0 = -1$  and for  $k \geq 1$  let

$$N_{2k-1} = \inf\{m > N_{2k-2}, X_m \leq a\}$$

$$N_{2k} = \inf\{m > N_{2k-1}, X_m \geq b\}$$

The  $N_j$  are stopping times so

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

defines a predictable sequence.



**Proof**

Notice

$$\{N_{2k-1} = n\} = \bigcup_{0 \leq m \leq n-1} \{N_{2k-2} = m\} \cap \left( \bigcap_{n-1-m \geq k \geq 0} \{X_{m+k} > a\} \right) \cap \{X_n \leq a\}$$

and

$$\{N_{2k} = n\} = \bigcup_{0 \leq m \leq n-1} \{N_{2k-1} = m\} \cap \left( \bigcap_{n-1-m \geq k \geq 0} \{X_{m+k} < b\} \right) \cap \{X_n \geq b\}$$

and hence  $N_{2k-1}, N_{2k}$  are stopping times by induction.

And notice

$$\{N_{2k-1} < m \leq N_{2k} \text{ for some } k\} = \bigcup_{k \geq 0} \{N_{2k-1} \leq m-1\} \cap \{N_{2k} \geq m\} \in \mathcal{F}_{m-1}$$

and hence  $H_m$  is predictable.

**Theorem 1.16**

(Upcoming inequality) If  $X_m, m \geq 0$ , is a submartingale then

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

where  $U_n = \sup\{k, N_{2k} \leq n\}$ .



**Proof**

Here we assume  $Y_m = a + (X_m - a)^+$  and we have

$$(b - a)U_n \leq (H \cdot Y)_n$$

let  $K_m = 1 - H_m$  and then we know that  $(K \cdot X)_n$  is a submartingale and then

$$E(K \cdot X)_n \geq E(K \cdot X)_0 = 0$$

so we know

$$E(H \cdot Y)_n \leq E(Y_n - Y_0) = E(X_n - a)^+ - E(X_0 - a)^+$$

since  $Y_n - Y_0 = (H \cdot Y)_n + (K \cdot Y)_n$

### Theorem 1.17

(Martingale convergence theorem) If  $X_n$  is a submartingale with  $\sup EX_n^+ < \infty$  then as  $n \rightarrow \infty$ ,  $X_n$  converges a.s. to a limit  $X$  with  $E|X| < \infty$ .



### Proof

We know  $(X - a)^+ \leq X^+ + |a|$ , then we know

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

so  $\sup EX_n^+$  will imply that  $EU < \infty$  where  $U = \lim U_n$  and hence for all rational  $a, b$ , we know

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

and hence  $\lim X_n$  exists a.s. and  $EX^+ \leq \liminf EX_n^+ < \infty$  and hence  $X < +\infty$  a.s. and notice

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

and hence  $EX^- \leq \liminf EX_n^- \leq \liminf EX_n^+ - EX_0 < \infty$  therefore  $E|X| < \infty$ .

### Theorem 1.18

If  $X_n \geq 0$  is a supermartingale then as  $n \rightarrow \infty$ ,  $X_n \rightarrow X$  a.s. and  $EX \leq EX_0$ .



### Proof

Let  $Y_n = -X_n$  and hence a submartingale with  $EY_n^+ = 0$ , then we know  $X_n \rightarrow X$  a.s. and we also have

$$EX \leq \liminf EX_n^+ \leq EX_0$$

### Proposition 1.2

The theorem 1.18. provide a method to show that a.s. convergence does not guarantee convergence in  $L^1$ .



### Proof

Let  $S_n$  be a symmetric simple random walk with  $S_0 = 1$  and  $P(\xi_i = 1) = P(\xi_i = -1) = \frac{1}{2}$ , let  $N = \inf\{n : S_n = 0\}$  and  $X_n = S_{N \wedge n}$ . Then we know  $X_n$  nonnegative and  $EX_n = EX_0 = 1$  since  $X_n$  is a martingale, then we know  $X_n \rightarrow X$  where  $X$  is some r.v. and hence  $X = 0$ , because there is no way to converge to others and hence  $X_n$  do not converge to  $X$  in  $L^1$ .

### Proposition 1.3

Convergence in probability do not guarantee convergence a.s.



### Proof

Let  $X_0 = 0$  and  $P(X_k = 1|X_{k-1} = 0) = P(X_k = -1|X_{k-1} = 0) = \frac{1}{2k}$ ,  $P(X_k = 0|X_{k-1} = 0) = 1 - \frac{1}{k}$  and  $P(X_k = kX_{k-1}|X_{k-1} \neq 0) = \frac{1}{k}$ ,  $P(X_k = 0|X_{k-1} \neq 0) = 1 - \frac{1}{k}$ , then we know  $X_k \rightarrow 0$  in probability, but  $P(X_k = 0, k \geq K)$  and it picks discrete values and hence  $X_k$  can not converge to 0 a.s.

**Theorem 1.19**

Let  $X_1, X_2, \dots$  be a martingale with  $|X_{n+1} - X_n| \leq M < \infty$ . Let

$$C = \{\lim X_n \text{ exists and is finite}\}$$

$$D = \{\limsup X_n = +\infty \text{ and } \liminf X_n = -\infty\}$$

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



**Proof** We may assume that  $X_0 = 0$  and then for  $K \geq 0$  denote

$$N = \inf\{n, X_n \leq -K\}$$

then we know  $X_{n \wedge N}$  is a martingale since

$$E(X_{(n+1) \wedge N} | \mathcal{F}_n) = E(X_{n+1} \chi_{N \geq n+1} + X_N \chi_{N \leq n} | \mathcal{F}_n) = X_N \chi_{N \leq n} + X_n \chi_{N \geq n+1} = X_{n \wedge N}$$

and  $X_{n \wedge N} \geq -K - M$  and hence  $X_{n \wedge N} + K + M \geq 0$  and we may know  $X_{n \wedge N}$  will converges to  $X$  a.s. with  $X$  finite. So if  $\liminf X_n > -\infty$ , then we know there exists  $K$  large enough such that  $N = \infty$  and hence  $X_n$  will converges to a finite limit on  $\{\liminf X_n > -\infty\}$ . For  $\limsup X_n$  consider  $N = \inf\{x, X_n \geq K\}$  with  $K + M - X_{n \wedge N}$  will converges and the  $\lim X_n$  will exists and be finite on  $\{\limsup X_n < +\infty\}$  and hence the conclusion holds.

**Theorem 1.20**

(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** If so we know

$$E(X_{n+1} | \mathcal{F}_n) = M_n + A_{n+1} = X_n - A_n + A_{n+1}$$

and hence set

$$A_n = \sum_{k=1}^n (E(X_k | \mathcal{F}_{k-1}) - X_{k-1})$$

and

$$M_k = X_k - A_k$$

then it is easy to check  $A_n$  is predictable increasing sequence and

$$E(M_{n+1} | \mathcal{F}_n) = E(X_{n+1} - \sum_{k=1}^{n+1} (E(X_k | \mathcal{F}_{k-1}) - X_{k-1}) | \mathcal{F}_n) = X_n - A_n = M_n$$

**Theorem 1.21**

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

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



**Proof** We know

$$\sum_{i=1}^{\infty} \chi_{B_i} = \infty$$

on  $\{B_n, i.o.\}$  and we know

$$\chi_{B_n} = M_n + \sum_{k=1}^n (E(\chi_{B_k} | \mathcal{F}_{k-1}) - \chi_{B_{k-1}})$$

and hence

$$M_n = \sum_{i=1}^n \chi_{B_i} - \sum_{i=1}^n E(\chi_{B_i} | \mathcal{F}_{i-1})$$



is a martingale. Then we know

$$EM_n = EX_0 < \infty$$

which means  $M_n$  is a martingale with bounded increments and we know

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

on both part of  $\Omega$ .

**Example 1.4** (Polya's Urn Scheme) An urn contains  $r$  red and  $g$  green balls. At each time we draw a ball out, then replact it with  $c$  balls with the same color. Let  $X_n$  be the fraction of green balls after the  $n^{\text{th}}$  draw.

**Proof**

$X_n$  is a martingale because assume  $\mathcal{F}_n$  is consisting by  $E_{i,j} = \{\text{There are } i \text{ green balls and } j \text{ red balls in the urn.}\}$  and it suffices to show that

$$\frac{j}{i+j} P(E_{i,j}) = \int_{E_{i,j}} E(X_{n+1} | \mathcal{F}_n)$$

where we know

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

and the equality is easy to be checked. Since  $X_n \geq 0$ , then we know  $X_n$  will converges to  $X$ .

#### Theorem 1.22

Assume  $\mu$  is a finite measure and  $\nu$  a probability measure on  $(\omega, \mathcal{F})$  with  $\mathcal{F}_n \uparrow \mathcal{F}$ , i.e.  $\sigma(\bigcup \mathcal{F}_n) = \mathcal{F}$  and  $\mu_n, \nu_n$  are the restrictions on  $\mathcal{F}_n$  of  $\mu, \nu$ . Suppose  $\mu_n \leq \nu_n$  for all  $n$ . Let  $X_n = \frac{d\mu_n}{d\nu_n}$  and let  $X = \limsup X_n$ , then

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

for any  $A \in \mathcal{F}$ .



**Proof**

We should show a lemma at first.

#### Lemma 1.3

$X_n$  defined on  $(\Omega, \mathcal{F}, \nu)$  is a martingale w.r.t.  $\mathcal{F}_n$ .



**Proof**

For any  $A \in \mathcal{F}_n$ , we know

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

and which means  $\int_A X_n d\nu = \int_A X_{n+1} d\nu$  for any  $A \in \mathcal{F}_n$ .

Now let's come back to the proof of the original theorem.

Now we know  $X_n$  is a nonnegative martingale on  $(\Omega, \mathcal{F}, \nu)$  and hence  $X_n \rightarrow X$   $\nu$ -a.s. Without loss of the generality, we may assume  $\mu$  is a probability measure and let  $\rho = (\mu + \nu)/2$ , then we know  $\mu \ll \nu \ll \rho$  and similarly define  $\rho_n$  and  $Y_n = d\mu_n/d\rho_n$ ,  $Z_n = d\nu_n/d\rho_n$  and  $Y_n + Z_n = 2$ ,  $Y_n, Z_n \geq 0$   $\rho_n$ -a.s. By the lemma, we will know that  $Y_n, Z_n$  are bounded martingales and we may assume they have limits  $Y, Z$ .

Notice for  $A \in \mathcal{F}_n$ , we have

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

by the DCT and hence  $\mu(A) = \int_A Y d\rho$  for all  $A \in \bigcup_m \mathcal{F}_m$  and we will know  $\mu(A) = \int_A Y d\rho$  for  $A \in \mathcal{F}$  by the  $\pi - \lambda$  theorem and hence  $Y = d\mu/d\rho$ , then we will know  $Z = d\nu/d\rho$ . Then notice

$$0 = \int_{\{Z_n=0\}} Z_n d\rho_n = \nu_n(\{Z_n=0\})$$

and hence  $\int_{Z_n=0} Y_n d\rho_n = \mu_n(\{Z_n=0\}) = 0$ , which means  $Y_n = 0$   $\rho$ -a.s. on  $\{Z_n=0\}$  which means  $Z_n > 0$  a.s. since

$\{Y_n = Z_n = 0\}$  is  $\rho$ -null. Then we know  $X_n = Y_n/Z_n$   $\rho$ -a.s. and hence  $X = Y/Z$   $\rho$ -a.s. and hence  $\nu$ -a.s.

Let  $W = (1/Z)\chi_{Z>0}$  and then  $1 = ZW + \chi_{Z=0}$  and we have


$$\mu(A) = \int_A YW Z d\rho + \int_A \chi_{Z=0} Y d\rho = \int_A X d\nu + \mu(A \cap \{X = \infty\})$$

since  $\nu(\{Z = 0\}) = \int_{\{Z=0\}} Z d\rho = 0$  and  $\{X = \infty\} = \{Z = 0\}$   $\rho$ -a.s. and hence  $\mu$ -a.s.


### Definition 1.9

Let  $\xi_i^n, i, n \geq 1$  be i.i.d. nonnegative integer-valued r.v.s and define

$$Z_{n+1} = \begin{cases} \xi_1^{n+1} + \dots + \xi_{Z_n}^{n+1} & Z_n > 0 \\ 0 & Z_n = 0 \end{cases}$$

where  $Z_0 = 1$  and  $Z_n$  is called a Galton – Watson process,  $p_k = P(\xi_i^n = k)$  is called the offspring distribution. 


### Lemma 1.4

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

**Proof** We know

$$E(Z_{n+1}/\mu^{n+1} | \mathcal{F}_n) = E\left(\sum \chi_{Z_n=k} \sum_{i=1}^k \xi_i^{n+1} / \mu^{n+1} \mid \mathcal{F}_n\right) = k \chi_{Z_n=k} / \mu^n = Z_n / \mu^n$$

### Theorem 1.23

If  $\mu < 1$  then  $Z_n = 0$  for all  $n$  sufficiently large, so  $Z_n/\mu^n \rightarrow 0$ . 

**Proof**

$E(Z_n/\mu^n) = E(Z_0) = 1$  and hence


$$P(Z_n > 0) \leq \mu^n$$

and hence

$$P(Z_n > 0 \text{ i.o.}) = 0$$

by the Borel-Cantelli's theorem, which means  $Z_n = 0$  for all  $n$  sufficiently large almost surely.


### Theorem 1.24

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


**Proof**

$2P(Z_n > 1) \leq \mu^n$  and hence  $Z_n \leq 1$  for all  $n$  sufficiently large almost surely, and the  $Z_n = 0$  for all  $n$  sufficiently large will not happen iff  $Z_n = 1$  for all  $n$  sufficiently large, which owns the probability of 0 and hence the conclusion holds.

### Definition 1.10

For  $s \in [0, 1]$ , let  $\phi(s) = \sum_{k \geq 0} p_k s^k$  and  $\phi$  is called the generating function for the offspring distribution  $p_k$ . 

### Theorem 1.25

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

**Proof**

Firstly let us show the existence. We can calculate

$$\phi'(s) = \sum k p_k s^{k-1}$$

by some methods in real analysis and hence  $\phi'(s) > h + \epsilon$  for some  $\epsilon > 0$  near 1 and hence there have to be a point in

$[0, 1)$  such that  $\phi(\rho) = \rho$  since  $\phi(0) \geq 0$ . And  $\phi'$  is increasing strictly on  $[0, 1)$  guaranteeing that the point is unique.

Then consider  $\theta_m = P(Z_m = 0)$ , then  $\theta_m = \phi(\theta_{m-1})$  which can be implied by consider  $Z_1 = k$  separately.

Then notice  $\theta_0 = 0$  and then  $\theta_m \leq \rho$  may imply that  $\theta_{m+1} = \phi(\theta_m) \leq \phi(\rho) \leq \rho$  and hence  $\phi(\theta_m) \geq \theta_m$ , which means  $\theta_m$  is increasing, then we know  $\theta_m \uparrow \rho$ .

### Theorem 1.26

If  $X_n$  is a submartingale and  $N$  is a stopping time with  $P(N \leq k) = 1$ , then

$$EX_0 \leq EX_N \leq EX_k$$



### Proof

We know  $X_{N \wedge n}$  is a submartingale since

$$E(X_{N \wedge (n+1)} | \mathcal{F}_n) = E(X_{n+1} \chi_{N \geq n+1} + X_N \chi_{N \leq n} | \mathcal{F}_n) \leq X_N \chi_{N \leq n} + X_n \chi_{N \geq n+1} = X_{N \wedge n}$$

and hence

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

and let  $K_n = \chi_{N < n}$  and we know  $K_n$  is predictable, then we know  $(K \cdot X)_n$  is a submartingale since

$$E((K \cdot X)_{n+1} | \mathcal{F}_n) = E\left(\sum_{m=1}^{n+1} K_m (X_m - X_{m-1}) | \mathcal{F}_n\right) = K_{n+1} E(X_{n+1} - X_n | \mathcal{F}_n) + (K \cdot X)_n \leq (K \cdot X)_n$$

since  $K$  is postive and we know

$$(K \cdot X)_n = X_n - X_{N \wedge n}$$

and we know

$$E(X_k - X_N) = E((K \cdot X)_k) \geq E((K \cdot X)_0) = 0$$

### Theorem 1.27

(Doob's inequality) Let  $X_m$  be a submartingale,

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

$\lambda > 0$  and  $A = \{\bar{X}_n \geq \lambda\}$ . Then

$$\lambda P(A) \leq EX_n \chi_A \leq EX_n^+$$



### Proof

Let  $N = \inf\{m, X_m \geq \lambda \text{ or } m = n\}$ , then  $X_N \geq \lambda$  on  $A$ , then

$$\lambda P(A) \leq EX_N \chi_A \leq EX_n \chi_A$$

by the theorem 1.26. and then second inequality is obvious.

2024/03/11