

# **Probability theory II Facts**

**Taeyoung Chang**

Textbook : Rick Durrett 「 Probability : Theory and Examples 」 5th edition

Last Update : October 15, 2021

# Contents

1	Conditional Expectation	2
2	Martingales	6

# 1 Conditional Expectation

- Projection Thm for Hilbert Space

- If  $E$  is a Hilbert space and  $M \subset E$  is closed and convex, then for any  $y \in E$ ,  
 $\exists$  a unique  $w \in M$  s.t.  $\|y - w\| = d(y, M) := \inf\{\|y - v\| : v \in M\}$ .  
Denote it as  $w = \text{proj}_M y$  i.e.  $w$  is a projection of  $y$  onto  $M$ .
- If  $E$  is a Hilbert space and  $M \subset E$  is a closed vector subspace, then for any  $y \in E$ ,
  - $\exists$  a unique decomposition  $y = w + v$  with  $w = \text{proj}_M y \in M$  and  $v \in M^\perp$
  - For  $w \in M$ ,  $w = \text{proj}_M y \Leftrightarrow \langle y - w, z \rangle = 0 \quad \forall z \in M$

\*  $\mathcal{L}^2 := \{ \text{Random Variable } X : E(X^2) = \int X^2 dP < \infty \}$

✓ If  $X \in \mathcal{L}^2$  then  $E|X| < \infty$  i.e. every element of  $\mathcal{L}^2$  is integrable.

★ Trick :  $|X| \leq X^2 + \frac{1}{4}$

✓  $\mathcal{L}^2$  is a vector space

★ Trick : inequality  $(aX + bY)^2 \leq 2(a^2X^2 + b^2Y^2)$

- $\mathcal{L}^2$  is a Hilbert space with inner product  $\langle X, Y \rangle = E(XY)$

★ Trick : Cauchy seq. having a subseq. converging to a point converges to the point.

- Lemma for proving  $\mathcal{L}^2$  is a complete normed space.

- If  $\{X_n\} \subset \mathcal{L}^2$  and  $\|X_n - X_{n+1}\| \leq 2^{-n} \quad \forall n \in \mathbb{N}$  then  $\exists X \in \mathcal{L}^2$  s.t.  $X_n \rightarrow X$  a.s. and  $\|X_n - X\| \rightarrow 0$  i.e.  $X_n \rightarrow X$  in  $\mathcal{L}^2$ .

★ Lemma : If a random variable  $Z$  satisfies  $Z \geq 0$  and  $E(Z) < \infty$  then  $Z < \infty$  a.s.

\* For  $X \in \mathcal{L}^2$ ,  $\mathcal{L}^2(X) := \{g(X) \mid g : \mathbb{R} \rightarrow \mathbb{R} \text{ is a Borel function, } E[(g(X))^2] < \infty\}$

✓ For  $X \in \mathcal{L}^2$ ,  $\mathcal{L}^2(X)$  is a vector subspace of  $\mathcal{L}^2$ .

- For  $X \in \mathcal{L}^2$ ,  $\mathcal{L}^2(X)$  is a closed vector subspace of  $\mathcal{L}^2$  so that  $\mathcal{L}^2(X)$  is also a Hilbert space.

\* Geometric definition for conditional expectation

- For  $X, Y \in \mathcal{L}^2$ , define  $E[Y|X] = \text{Proj}_{\mathcal{L}^2(X)} Y$
- $E[Y|X] = g(X)$  a.s. for some Borel function  $g$
- $\|Y - E[Y|X]\| = \min_{h(X) \in \mathcal{L}^2(X)} \|Y - h(X)\|$   
i.e.  $E[(Y - E[Y|X])^2] \leq E[(Y - h(X))^2] \quad \forall h(X) \in \mathcal{L}^2$
- For  $g(X) \in \mathcal{L}^2(X)$ ,  $g(X) = E[Y|X] \Leftrightarrow \langle Y - g(X), h(X) \rangle = 0 \quad \forall h(X) \in \mathcal{L}^2$   
 $\Leftrightarrow E[(Y - g(X))h(X)] = 0 \quad \forall h(X) \in \mathcal{L}^2$

- Elementary properties of conditional expectation from geometric definition

- If  $X, Y, Z \in \mathcal{L}^2$  then the followings are true.
  - $E[c|X] = c$  a.s.  $\forall c \in \mathbb{R}$
  - $E[\alpha Y + \beta Z|X] = \alpha E[Y|X] + \beta E[Z|X] \quad \forall \alpha, \beta \in \mathbb{R}$
  - $E[Y|X] = E[Y]$  if  $X$  and  $Y$  are independent.

iv.  $E[g(X)Y|X] = g(X)E[Y|X]$  if  $g$  satisfies  $g(X) \in \mathcal{L}^2(X)$  and  $\sup_x |g(x)| < \infty$

v.  $E[E[Y|X]] = E[Y]$

✓ In fact, the additional assumption about boundedness of  $g$  in (iv) is not necessary. We will see later.

- Extending the definition from  $\mathcal{L}^2$  to all integrable functions

$$E[\{Y - E[Y|X]\}I(X \in A)] = 0 \quad \forall A \in \mathcal{B}(\mathbb{R}) \quad \because I(X \in A) \in \mathcal{L}^2(X)$$

$$\int_{(X \in A)} Y dP = \int_{(X \in A)} E[Y|X] dP \quad \forall A \in \mathcal{B}(\mathbb{R})$$

$$\int_B Y dP = \int_B E[Y|X] dP \quad \forall B \in \sigma(X)$$

- $E[Y|X] \in \sigma(X)$  and  $\int_B Y dP = \int_B E[Y|X] dP \quad \forall B \in \sigma(X)$ . Such r.v. is unique in the sense that if any r.v.  $Z$  satisfies  $Z \in \sigma(X)$  and  $\int_B Y dP = \int_B Z dP \quad \forall B \in \sigma(X)$  then  $Z = E[Y|X]$  a.s. provided  $E|Y| < \infty$
- From the theory on  $\mathcal{L}^2$  space, we get geometric understanding about conditional expectation. But now, from the equation above, we can guess that definition for conditional expectation may be extended to all integrable random variables.

- Proof for the uniqueness mentioned above

- $(\Omega, \mathcal{F}, P)$  : a prob. space.  $Y \in \mathcal{F}$  and  $E|Y| < \infty$ .  $\mathcal{G} \subset \mathcal{F}$  is a sub  $\sigma$ -field. If  $X$  is a random variable satisfying (a)  $X \in \mathcal{G}$  (b)  $\int_A Y dP = \int_A X dP \quad \forall A \in \mathcal{G}$  then
  - $X$  is integrable
  - Such  $X$  is unique in the sense that if there is another  $X'$  then  $X = X'$  a.s.
    - ★ Trick : For any r.v.  $Z$ ,  $(Z > 0) = \bigcup_{\epsilon > 0} (Z \geq \epsilon) = \bigcup_{n \in \mathbb{N}} (Z > \frac{1}{n})$
    - ★ Lemma : For any  $\mathcal{F}$ -measurable and integrable  $X$  and  $Y$ , if  $\int_A X dP = \int_A Y dP \quad \forall A \in \mathcal{F}$  then  $X = Y$  a.s.

- Radon-Nikodym Thm

- If  $\mu, \nu$  are  $\sigma$ -finite measures on  $(\Omega, \mathcal{F})$  and  $\nu \ll \mu$  ( $\mu(A) = 0 \Rightarrow \nu(A) = 0 \quad \forall A \in \mathcal{F}$ ) then  $\exists$  a  $\mathcal{F}$ -measurable nonnegative function  $g$  s.t.  $\nu(A) = \int_A g d\mu \quad \forall A \in \mathcal{F}$ . The function  $g$  is unique in the sense that if  $h$  is another such function then  $g = h$   $\mu$ -a.e.

- \* Definition of conditional expectation

- $(\Omega, \mathcal{F}_0, P)$  : a prob. space.  $\mathcal{F} \subset \mathcal{F}_0$  : a sub  $\sigma$ -field.  
 $X$  is a random variable s.t.  $X \geq 0$ ,  $X \in \mathcal{F}_0$  and  $E|X| < \infty$ . Then  $\exists$  a unique r.v.  $Y$  s.t.  $Y \geq 0$ ,  $Y \in \mathcal{F}$  and  $\int_A X dP = \int_A Y dP \quad \forall A \in \mathcal{F}$ . Such  $Y$  is unique in the sense that if another  $Y'$  exists then  $Y = Y'$  a.s.
- $Y = E[X|\mathcal{F}]$  is said to be conditional expectation of  $X$  given  $\mathcal{F}$ 
  - ★ Applying Radon Nikodym thm to measures  $P|_{\mathcal{F}}$  and  $Q$  on  $(\Omega, \mathcal{F})$  where  $Q$  is defined by  $Q(A) = \int_A X dP \quad \forall A \in \mathcal{F}$ . Note that  $Q \ll P|_{\mathcal{F}}$  and  $Q$  is a finite measure.
- We can extend the definition to general integrable r.v.  $X$   
 $Y = E[X|\mathcal{F}]$  is a unique random variable s.t.  $Y \in \mathcal{F}$  and  $\int_A X dP = \int_A Y dP \quad \forall A \in \mathcal{F}$ .  
 $E[X|\mathcal{F}]$  is also integrable and the uniqueness is in the sense of a.s. equivalence relation.  
 $Y = E[X|\mathcal{F}]$  can be derived by  $Y = Y_1 - Y_2$  where  $Y_1 = E[X^+|\mathcal{F}]$  and  $Y_2 = E[X^-|\mathcal{F}]$

\* Conditional expectation given a random variable

–  $X$  : integrable r.v. For a random variable  $Y$ , define  $E[X|Y] := E[X|\sigma(Y)]$

✓  $Y$  need not be integrable.

✓ Since  $E[X|Y] \in \sigma(Y)$ ,  $E[X|Y] = g(Y)$  for some Borel function  $g$ . This coincides with the definition of conditional expectation in  $\mathcal{L}^2$  space.

\* Conditional probability

– For  $A \in \mathcal{F}_0$  and a sub  $\sigma$ -field  $\mathcal{F} \subset \mathcal{F}_0$ , define  $P(A|\mathcal{F}) := E[I_A|\mathcal{F}]$

– For  $A, B \in \mathcal{F}_0$ , define  $P(A|B) = P(A \cap B) / P(B)$

• Elementary properties of conditional expectation

–  $(\Omega, \mathcal{F}_0, P)$  : a prob. space.  $\mathcal{F} \subset \mathcal{F}_0$  : a sub  $\sigma$ -field.  $X, Y$  : integrable random variables

i.  $E[c|\mathcal{F}] = c$

ii.  $E[\psi(X)|\mathcal{F}] = \psi(X)$  given  $E|\psi(X)| < \infty$

iii. If  $\mathcal{F}$  is a trivial  $\sigma$ -field i.e.  $\mathcal{F} = \{\Omega, \emptyset\}$  then  $E[X|\mathcal{F}] = E[X]$

iv.  $\Omega = \bigcup_{i=1}^{\infty} \Omega_i$  is a partition of  $\Omega$  with  $\Omega_i \in \mathcal{F}_0$  and  $P(\Omega_i) > 0 \quad \forall i \in \mathbb{N}$   
 $\mathcal{F} = \sigma\{\Omega_1, \Omega_2, \dots\} = \{\bigcup_{j \in \kappa} \Omega_j : \kappa \subset \mathbb{N}\} \quad (\mathcal{F} \text{ is a } \sigma\text{-field}).$  Then we have

$$E[X|\mathcal{F}] = \sum_{i=1}^{\infty} a_i I_{\Omega_i} \quad \text{with} \quad a_i = \frac{E[X I_{\Omega_i}]}{P(\Omega_i)}$$

✓ For  $A \in \mathcal{F}_0$ ,  $P(A|\mathcal{F}) = P(A|\Omega_i) I_{\Omega_i}$

★ Lemma : If  $Z \in \mathcal{F}$  for such  $\mathcal{F}$ , then we can write  $Z = \sum_{i=1}^{\infty} c_i I_{\Omega_i}$  where  $c_i \in \mathbb{R}$

v.  $E[aX + bY|\mathcal{F}] = aE[X|\mathcal{F}] + bE[Y|\mathcal{F}] \quad \forall a, b \in \mathbb{R}$

vi.  $X \geq 0 \Rightarrow E[X|\mathcal{F}] \geq 0 \quad a.s.$

★ Lemma : If  $Z > 0$  on  $A$  with  $P(A) > 0$  then  $\int_A Z dP > 0$

vii.  $X \leq Y \Rightarrow E[X|\mathcal{F}] \leq E[Y|\mathcal{F}] \quad a.s.$

viii.  $|E[X|\mathcal{F}]| \leq E[|X||\mathcal{F}]$

□  $|X| \leq M$  for some  $M > 0 \Rightarrow |E[X|\mathcal{F}]| \leq M \quad a.s.$

ix.  $E[|X||\mathcal{F}] = 0 \Rightarrow X = 0 \quad a.s.$

x.  $E[E[X|\mathcal{F}]] = E[X]$

•  $X, Y$  : integrable r.v's where  $X \perp\!\!\!\perp Y$ .  $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$  Borel measurable s.t.  $E|\psi(X, Y)| < \infty$   
 Define  $g : \mathbb{R} \rightarrow \mathbb{R}$  by  $g(x) = E[\psi(x, Y)] \quad \forall x \in \mathbb{R}$ . Then  $E[\psi(X, Y)|X] = g(X)$

✓  $g(x) = E[\psi(x, Y)] = \int \psi(x, Y) dP = \int_{\mathbb{R}} \psi(x, y) dPY^{-1}(y) = \int_{\mathbb{R}} \psi_x(y) d\mu_Y(y) \quad \forall x \in \mathbb{R}$   
 By Fubini thm in real analysis course, it is shown that  $g$  is Borel measurable & integrable.

• Conditional expectation and convergence

–  $(\Omega, \mathcal{F}_0, P)$  : a probability space.  $\mathcal{F} \subset \mathcal{F}_0$  : a sub  $\sigma$ -field

i. (MCT) If  $X_n \geq 0$  and  $X_n \nearrow X$  a.s. with  $E(X) < \infty$  then  $E[X_n|\mathcal{F}] \nearrow E[X|\mathcal{F}]$  a.s.

□ If  $Y_n \searrow Y$  a.s. with  $E|Y_1|, E|Y| < \infty$  then  $E[Y_n|\mathcal{F}] \searrow E[Y|\mathcal{F}]$  a.s.

ii. (DCT) If  $|X_n| \leq Y$ ,  $E(Y) < \infty$  and  $X_n \rightarrow X$  a.s. then  $E[X_n|\mathcal{F}] \rightarrow E[X|\mathcal{F}]$  a.s.

- iii. (Fatou's lemma) If  $X_n \geq 0$  and  $X_n \rightarrow X$  a.s. with  $E(X_n) < \infty$ ,  $E(X) < \infty$  then  $E[X|\mathcal{F}] \leq \liminf E[X_n|\mathcal{F}]$
- iv. (Continuity from below)  $\{B_n\} \subset_{seq} \mathcal{F}_0$  s.t.  $B_n \subset B_{n+1} \quad \forall n \in \mathbb{N}$ .  $B := \bigcup_n B_n$   
Then  $P(B_n|\mathcal{F}) \nearrow P(B|\mathcal{F})$
- v. (Countable additivity) If  $\{C_n\} \subset_{seq} \mathcal{F}_0$  is mutually disjoint then  $P(\bigcup_n C_n|\mathcal{F}) = \sum_n P(C_n|\mathcal{F})$

- Essential inequalities

- i. (Markov)  $P(|X| \geq c|\mathcal{F}) \leq \frac{1}{c}E[|X||\mathcal{F}] \quad \forall c > 0$
- ii. (Jensen) If  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is convex then  $\phi(E[X|\mathcal{F}]) \leq E[\phi(X)|\mathcal{F}]$  a.s.  
★ Trick : For each  $x \in \mathbb{R}$  and convex function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$ , we have  $\phi(x) = \sup\{ax + b : (a, b) \in S\}$  where  $S = \{(a, b) \in \mathbb{R}^2 : ax + b \leq \phi(x) \quad \forall x \in \mathbb{R}\}$
- iii. (Cauchy-Schwarz) For  $X, Y \in \mathcal{L}^2$ , we have  $E^2[XY|\mathcal{F}] \leq E[X^2|\mathcal{F}]E[Y^2|\mathcal{F}]$  a.s.

- Smoothing property of conditional expectation

- i. If  $X \in \mathcal{F}$ ,  $E|Y| < \infty$ , and  $E|XY| < \infty$  then  $E[XY|\mathcal{F}] = XE[Y|\mathcal{F}]$  a.s.  
 $\checkmark$   $E|X| < \infty$  assumption is not required.
- If  $X \in \mathcal{F}$  and  $E|X| < \infty$  then  $E[X|\mathcal{F}] = X$  a.s.
- ii. If  $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_0$  are sub  $\sigma$ -fields and  $E|X| < \infty$  then
  - (a)  $E[E(X|\mathcal{F}_1)|\mathcal{F}_2] = E[X|\mathcal{F}_1]$
  - (b)  $E[E(X|\mathcal{F}_2)|\mathcal{F}_1] = E[X|\mathcal{F}_1]$
- ★ Lemma : If  $\mathcal{F}_1 \subset \mathcal{F}_2$  then  $Y \in \mathcal{F}_1 \Rightarrow Y \in \mathcal{F}_2$   
 $\checkmark$  In short, “the smaller wins”. In view of information, it is similar to projection onto vector subspaces  $S_1 \subset S_2 \subset S$  where  $Proj_{S_1} Proj_{S_2} = Proj_{S_2} Proj_{S_1} = Proj_{S_1}$

- Def. of conditional expectation by Radon-Nikodym derivative agrees with def. in  $\mathcal{L}^2$  space.

- If  $E(X^2) < \infty$  then for  $\mathcal{C} = \{Y : Y \in \mathcal{F}, E(Y^2) < \infty\}$ ,  
 $E[\{X - E[X|\mathcal{F}]\}^2] = \inf_{Y \in \mathcal{C}} E[\{X - Y\}^2]$  and  $E[X|\mathcal{F}] = \arg \min_{Y \in \mathcal{C}} E[\{X - Y\}^2]$
- ★ Lemma : If  $X \in \mathcal{L}^2$  then  $E[X|\mathcal{F}] \in \mathcal{L}^2$

- \* Independence of a random variable and a  $\sigma$ -field

- A random variable  $X$  and a  $\sigma$ -field  $\mathcal{F}$  are said to be independent if  $\sigma(X)$  and  $\mathcal{F}$  are independent

- If an integrable random variable  $X$  and a  $\sigma$ -field  $\mathcal{F}$  are independent then  $E[X|\mathcal{F}] = E[X]$

- Two extreme cases of conditional expectations w.r.t information

- Perfect information : If  $X \in \mathcal{F}$  then  $E[X|\mathcal{F}] = X$
- No information : If  $X \perp \mathcal{F}$  then  $E[X|\mathcal{F}] = E[X]$

- \* Conditional variance

$$Var(X|\mathcal{F}) := E[\{X - E[X|\mathcal{F}]\}^2|\mathcal{F}] = E[X^2|\mathcal{F}] - E^2[X|\mathcal{F}]$$

Conditional variance is defined for  $X \in \mathcal{L}^2$

## 2 Martingales

\* Definition needed for martingales

- Given a probability space  $(\Omega, \mathcal{F}, P)$ , increasing sequence of sub  $\sigma$ -fields  $\{\mathcal{F}_n\}_{n=0}^\infty$  is called a filtration.
- A random sequence  $\{X_n\}_{n=0}^\infty$  is said to be adapted to  $\{\mathcal{F}_n\}$  if  $X_n \in \mathcal{F}_n \quad \forall n \in \mathbb{N} \cup \{0\}$

\* Definition of martingale and their cousins

- $\{X_n\}_{n=0}^\infty$  : a random sequence.  $\{\mathcal{F}_n\}_{n=0}^\infty$  : a filtration. Assume  $E|X_n| < \infty \quad \forall n \in \mathbb{N} \cup \{0\}$  and  $\{X_n\}$  is adapted to  $\{\mathcal{F}_n\}$ . Then  $\{X_n\}$  is said to be a martingale (w.r.t  $\{\mathcal{F}_n\}$ ) if  $E[X_{n+1}|\mathcal{F}_n] = X_n \quad \forall n \in \mathbb{N} \cup \{0\}$
  - $\{X_n\}$  is said to be a submartingale (w.r.t  $\{\mathcal{F}_n\}$ ) if  $E[X_{n+1}|\mathcal{F}_n] \geq X_n \quad \forall n \in \mathbb{N} \cup \{0\}$
  - $\{X_n\}$  is said to be a supermartingale (w.r.t  $\{\mathcal{F}_n\}$ ) if  $E[X_{n+1}|\mathcal{F}_n] \leq X_n \quad \forall n \in \mathbb{N} \cup \{0\}$
- ✓ These are abbreviated to ‘mtg’, ‘submtg’, ‘supermtg’ respectively.

• Examples of martingales

- i.  $\{\xi_n\}_n$  i.i.d with  $E(\xi_1) = 0$ .  $X_0 = 0$ .  $X_n = \xi_1 + \dots + \xi_n$  and  $\mathcal{F}_0 = \{\phi, \Omega\}$ .  
 $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$ . Then  $\{X_n\}$  is a martingale w.r.t  $\{\mathcal{F}_n\}$   
 ★ Trick :  $E[Z]$  is finite  $\Leftrightarrow Z$  is integrable. ( $\because$  the definition of expectation)
- ii. Adding assumption  $Var(\xi_1) = \sigma^2 < \infty$  to i. above.  
 Then  $\{X_n - n\sigma^2\}$  is a martingale w.r.t  $\{\mathcal{F}_n\}$
- iii.  $\{\varepsilon_n\}_n$  i.i.d  $\sim (0, 1)$ .  $X_0 = 0$ .  $X_{n+1} = X_n + h(X_n)\varepsilon_{n+1}$  with  $h : \mathbb{R} \rightarrow \mathbb{R}$  Borel function s.t.  
 $E|h(X_n)| < \infty \quad \forall n \in \mathbb{N} \cup \{0\}$  and  $\mathcal{F}_0 = \{\phi, \Omega\}$ .  $\mathcal{F}_n = \sigma(\varepsilon_1, \dots, \varepsilon_n)$   
 Then  $\{X_n\}$  is a martingale w.r.t  $\{\mathcal{F}_n\}$
- iv.  $\{\varepsilon_n\}_n$  i.i.d  $\sim (0, 1)$ .  $Y_0 = 0$ .  $Y_{n+1} = \phi(Y_n)\varepsilon_{n+1}$  with  $\phi(y) = w + \alpha y^2$  ( $w > 0, 0 \leq \alpha < 1$ )  
 and  $E[\phi(Y_n)] < \infty \quad \forall n \in \mathbb{N}$ . and  $\mathcal{F}_0 = \{\phi, \Omega\}$ .  $\mathcal{F}_n = \sigma(\varepsilon_1, \dots, \varepsilon_n)$ .  
 Let  $X_0 = 0$ .  $X_n = Y_1 + \dots + Y_n$ . Then  $\{X_n\}$  is a martingale w.r.t  $\{\mathcal{F}_n\}$   
 ✓ Such  $\{Y_n\}$  is called as ARCH (autoregressive conditional heteroskedasticity) process

• Elementary facts about Martingales

- Every martingale is a submartingale and a supermartingale
- If  $\{X_n\}$  is a submartingale then  $\{-X_n\}$  is a supermartingale  
 ✓ We develop theory about martingales often assuming submartingale since every martingale is submartingale and every supermartingale is negative version of submartingale
- If  $\{X_n\}$  is a martingale w.r.t  $\{\mathcal{F}_n\}$  then  $E[X_n|\mathcal{F}_m] = X_m$  whenever  $n \geq m$
- If  $\{X_n\}$  is a submartingale w.r.t  $\{\mathcal{F}_n\}$  then  $E[X_n|\mathcal{F}_m] \geq X_m$  whenever  $n \geq m$
- If  $\{X_n\}$  is a supermartingale w.r.t  $\{\mathcal{F}_n\}$  then  $E[X_n|\mathcal{F}_m] \leq X_m$  whenever  $n \geq m$
- If  $\{X_n\}$  is a martingale w.r.t  $\{\mathcal{F}_n\}$  then  $\{E[X_n]\}$  is constant.
- If  $\{X_n\}$  is a submartingale w.r.t  $\{\mathcal{F}_n\}$  then  $\{E[X_n]\}$  is increasing.
- If  $\{X_n\}$  is a supermartingale w.r.t  $\{\mathcal{F}_n\}$  then  $\{E[X_n]\}$  is decreasing.

- Convex transformation of martingale

- If  $\{X_n\}$  is a martingale w.r.t  $\{\mathcal{F}_n\}$  and  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is a convex function s.t.  $E|\phi(X_n)| < \infty \quad \forall n \in \mathbb{N}$  then  $\{\phi(X_n)\}$  is a submartingale w.r.t  $\{\mathcal{F}_n\}$
- If  $\{X_n\}$  is a submartingale w.r.t  $\{\mathcal{F}_n\}$  and  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is a convex and increasing function s.t.  $E|\phi(X_n)| < \infty \quad \forall n \in \mathbb{N}$  then  $\{\phi(X_n)\}$  is a submartingale w.r.t  $\{\mathcal{F}_n\}$
- If  $\{X_n\}$  is a supermartingale w.r.t  $\{\mathcal{F}_n\}$  and  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is a concave and increasing function s.t.  $E|\phi(X_n)| < \infty \quad \forall n \in \mathbb{N}$  then  $\{\phi(X_n)\}$  is a supermartingale w.r.t  $\{\mathcal{F}_n\}$
- (Ex) If  $\{X_n\}$  is a martingale and  $E[|X_n|^p] < \infty$  for some  $p \geq 1$ , then  $\{|X_n|^p\}$  is a submartingale
- (Ex) If  $\{X_n\}$  is a submartingale then for any  $a \in \mathbb{R}$ ,  $\{(X_n - a)^+\}$  is a submartingale
- (Ex) If  $\{X_n\}$  is a supermartingale then for any  $a \in \mathbb{R}$ ,  $\{X_n \wedge a\}$  is a supermartingale

- \* Predicable sequence and a process using it

- For a filtration  $\{\mathcal{F}_n\}_{n=0}^\infty$ , a random sequence  $\{H_n\}_{n=1}^\infty$  is said to be a predicable sequence (w.r.t  $\{\mathcal{F}_n\}$ ) if  $H_n \in \mathcal{F}_{n-1} \quad \forall n \in \mathbb{N}$
- ✓ A letter  $H$  stands for a ‘height’
- Suppose  $\{X_n\}$  is adapted to  $\{\mathcal{F}_n\}$ . For a predicable sequence  $\{H_n\}$  (w.r.t  $\{\mathcal{F}_n\}$ ), we define a process  $\{(H \cdot X)_n\}$  by

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

- ✓ Note that  $\{(H \cdot X)_n\}$  is adapted to  $\{\mathcal{F}_n\}$
- ✓ The definition above can be extended from  $\{(H \cdot X)_n\}_{n \in \mathbb{N}}$  to  $\{(H \cdot X)_n\}_{n \in \mathbb{N} \cup \{0\}}$  with additionally defining  $(H \cdot X)_0 = 0$ . Obviously  $(H \cdot X)_0 \in \mathcal{F}_0$ . For the following theorems using this process, we can regard it as  $\{(H \cdot X)_n\}_{n \in \mathbb{N} \cup \{0\}}$

- Elementary facts about martingale transform with predicable sequence

- Let  $\{X_n\}_{n=0}^\infty$  and  $\{H_n\}_{n=1}^\infty$  be a random sequence and  $\{H_n\}$  is a predicable sequence w.r.t. a filtration  $\{\mathcal{F}_n\}_{n=0}^\infty$ . Assume  $E|X_n H_n| < \infty$ ,  $E|X_{n-1} H_n| < \infty \quad \forall n \in \mathbb{N}$ 
  - i. If  $\{X_n\}$  is a martingale (w.r.t  $\{\mathcal{F}_n\}$ ) then  $\{(H \cdot X)_n\}$  is also a martingale
  - ii. If  $\{X_n\}$  is a submartingale (w.r.t  $\{\mathcal{F}_n\}$ ) and  $H_n \geq 0$  then  $\{(H \cdot X)_n\}$  is also a submartingale
  - iii. If  $\{X_n\}$  is a supermartingale (w.r.t  $\{\mathcal{F}_n\}$ ) and  $H_n \geq 0$  then  $\{(H \cdot X)_n\}$  is also a supermartingale
- ✓ The condition “ $E|X_n H_n| < \infty$ ,  $E|X_{n-1} H_n| < \infty \quad \forall n \in \mathbb{N}$ ” can be replaced with “For each  $n \in \mathbb{N}$ ,  $H_n$  is bounded”.

- \* Stopping time

- A (extended) random variable  $N$  taking values of  $\mathbb{N} \cup \{0, \infty\}$  is said to be a stopping time (w.r.t a filtration  $\{\mathcal{F}_n\}$ ) if an event  $(N = n) \in \mathcal{F}_n \quad \forall n \in \mathbb{N}$

$$\begin{aligned} (N \leq n) &= \bigcup_{j=0}^n (N = j) \in \mathcal{F}_n & (N > n) &= (N \leq n)^C \in \mathcal{F}_n \\ (N < n) &= \bigcup_{j=0}^{n-1} (N = j) \in \mathcal{F}_{n-1} & (N \geq n) &= (N < n)^C \in \mathcal{F}_{n-1} \end{aligned}$$



- $(N \geq n)$  is a  $\mathcal{F}_{n-1}$ -measurable event.  $I(N \geq n)$  is  $\mathcal{F}_{n-1}$ -measurable random variable. Hence,  $\{I(N \geq n)\}_n$  is a predictable sequence given  $N$  is a stopping time.

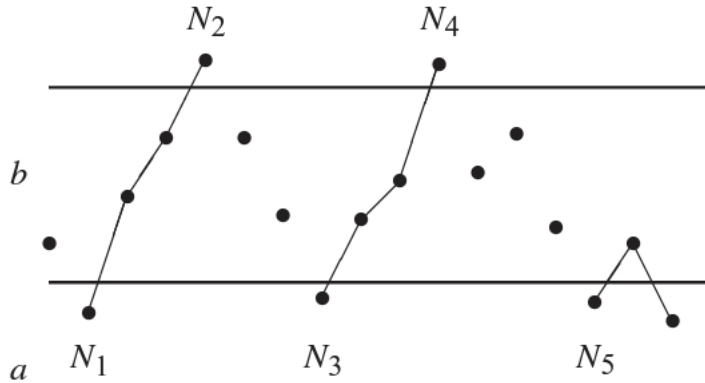
- Martingale stopped by stopping time

- Let  $\{X_n\}$  be a random sequence adapted to  $\{\mathcal{F}_n\}$ . Let  $N$  be a stopping time w.r.t  $\{\mathcal{F}_n\}$  and put  $H_n = I(N \geq n) \quad \forall n \in \mathbb{N}$ . Then  $(H \cdot X)_n = X_{N \wedge n} - X_0$ .
- The process  $\{X_{N \wedge n}\}_n$  is said to be a martingale stopped by stopping time  $N$ , provided  $\{X_n\}$  is a martingale.
  - ★ If  $\{X_n\}$  and  $\{Y_n\}$  are martingales (w.r.t.  $\{\mathcal{F}_n\}$ ) then  $\{X_n + Y_n\}$  is also a martingale. The same holds for submartingales and supermartingales too.
- If  $\{X_n\}$  is a martingale and  $N$  is a stopping time then  $\{X_{N \wedge n}\}$  is martingale.
- If  $\{X_n\}$  is a submartingale and  $N$  is a stopping time then  $\{X_{N \wedge n}\}$  is submartingale.
- If  $\{X_n\}$  is a supermartingale and  $N$  is a stopping time then  $\{X_{N \wedge n}\}$  is supermartingale.

- Stopping time and Upcrossing

- Suppose  $\{X_n\}_{n=0}^\infty$  is a submartingale w.r.t  $\{\mathcal{F}_n\}$ . Let  $a < b$ . Define  $N_j$ 's as below :

$$\begin{aligned}
 N_1 &= \inf\{m \geq 0 : X_m \leq a\} & N_2 &= \inf\{m > N_1 : X_m \geq b\} \\
 N_3 &= \inf\{m > N_2 : X_m \leq a\} & N_4 &= \inf\{m > N_3 : X_m \geq b\} \\
 &\vdots & &\vdots \\
 N_{2k-1} &= \inf\{m > N_{2k-2} : X_m \leq a\} & N_{2k} &= \inf\{m > N_{2k-1} : X_m \geq b\} \\
 &\vdots & &\vdots
 \end{aligned}$$



- Every  $N_j$  for  $j \in \mathbb{N}$  is stopping time w.r.t  $\{\mathcal{F}_n\}$ .  $N_1 < N_2 < N_3 \dots$  provided all  $N_j$ 's are finite. (It is possible that  $N_j = \infty$  provided it has a form of  $\inf(\emptyset)$ )
- ‘Upcrossing’ is a case where the submartingale  $\{X_n\}$  crosses from below  $a$  to above  $b$ .
- $\mathcal{U}_n := \sup\{k : N_{2k} \leq n\}$  is the number of upcrossings completed by time  $n$

- Upcrossing inequality

- Suppose  $\{X_n\}$  is a submartingale w.r.t  $\{\mathcal{F}_n\}$ . If stopping time  $N_j$  and the number of upcrossings  $\mathcal{U}_n$  are defined as above then

$$(b - a)E[\mathcal{U}_n] \leq E[(X_n - a)^+] - E[(X_0 - a)^+]$$

- Submartingale convergence theorem

- If  $\{X_n\}$  is a submartingale w.r.t  $\{\mathcal{F}_n\}$  with  $\sup_n E(X_n^+) < \infty$  then  $X_n \rightarrow X$  a.s. for some integrable random variable  $X$

- ★ Trick : If  $X_n \rightarrow X$  a.s. then  $X_n^+ \rightarrow X^+$  a.s. and  $X_n^- \rightarrow X^-$  a.s.

- ★ Lemma : If the number of upcrossings of  $[a, b]$  by submartingale  $\{X_n\}$  is finite for any  $a, b \in \mathbb{Q}$ , then  $\lim_n X_n$  exists. i.e.  $X_n$  converges to some r.v. almost surely.

- If  $\{X_n\}$  is a nonnegative supermartingale w.r.t  $\{\mathcal{F}_n\}$  then  $X_n \rightarrow X$  a.s. for some integrable random variable  $X$  s.t.  $E(X) \leq E(X_0)$

- Example of martingale which converges almost surely but not in  $L^1$

- $\{\xi_n\}_n$  i.i.d with  $P(\xi_1 = 1) = P(\xi_1 = -1) = 1/2$ . Let  $S_0 = 1$ ,  $S_n = S_{n-1} + \xi_n$  and  $\mathcal{F}_0 = \{\phi, \Omega\}$ ,  $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$ . Then  $\{S_n\}$  is a martingale w.r.t  $\{\mathcal{F}_n\}$

Let  $N = \inf\{n \in \mathbb{N} : S_n = 0\}$ . Then  $N$  is a stopping time.

$X_n := S_{N \wedge n}$  so that  $X_n = S_n$  if  $n < N$  and  $X_n = 0$  if  $n \geq N$ .  $\{X_n\}$  is a nonnegative integer valued martingale w.r.t  $\{\mathcal{F}_n\}$ .  $X_n \rightarrow 0$  a.s. but  $X_n \not\rightarrow 0$  in  $\mathcal{L}^1$ .

- If  $\{X_n\}_{n \in \mathbb{N} \cup \{0\}}$  is a negative submartingale w.r.t  $\{\mathcal{F}_n\}_{n \in \mathbb{N} \cup \{0\}}$  then so is  $\{X_n\}_{n \in \mathbb{N} \cup \{0, \infty\}}$  w.r.t  $\{\mathcal{F}_n\}_{n \in \mathbb{N} \cup \{0, \infty\}}$  where  $X_\infty = \lim_n X_n$  and  $\mathcal{F}_\infty = \sigma(\bigcup_{n=0}^\infty \mathcal{F}_n)$

- Doob's decomposition

- Any submartingale  $\{X_n\}$  can be written as  $X_n = M_n + A_n$  where  $\{M_n\}$  is a martingale and  $\{A_n\}$  is a predictable increasing sequence with  $A_0 = 0$ . Also, this expression is unique in the sense that if  $X_n = M'_n + A'_n$  is another expression then  $M_n = M'_n$  and  $A_n = A'_n$  a.s.

✓ The exact form of  $M_n, A_n$  for given  $X_n$  is  $A_n = A_{n-1} + E[X_n | \mathcal{F}_{n-1}] - X_{n-1} \quad \forall n \in \mathbb{N}$  and  $M_n = X_n - A_n \quad \forall n \in \mathbb{N} \cup \{0\}$  ( Since  $A_0 = 0$ ,  $A_n = \sum_{k=1}^n (E[X_k | \mathcal{F}_{k-1}] - X_{k-1})$  )

- Martingales with bounded increments either converge or oscillate between  $\infty$  and  $-\infty$

- Let  $\{X_n\}$  be a martingale with  $|X_n - X_{n-1}| \leq M < \infty \quad \forall n \in \mathbb{N}$  for some  $M > 0$ . Define disjoint subsets  $C, D \subset \Omega$  by

$$C = ( \lim_n X_n \text{ exists and } -\infty < \lim_n X_n < \infty )$$

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

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

- ★ Define “ $X_n \rightarrow X$  a.s. on  $B$ ” for measurable set  $B$  as  $P((X_n \rightarrow X) \cap B) = P(B)$

- ★ Trick :  $X_n \rightarrow X$  a.s. on  $B \Rightarrow X_n \rightarrow X$  a.s. on  $A$  whenever  $A \subset B$

- Conditional Borel-Cantelli second lemma

- Let  $\{\mathcal{F}_n\}_{n \in \mathbb{N} \cup \{0\}}$  be a filtration with  $\mathcal{F}_0 = \{\phi, \Omega\}$ . If  $A_n \in \mathcal{F}_n \quad \forall n \in \mathbb{N}$  then

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

- ★ Define “ $A = B$  a.s.” for measurable sets  $A$  and  $B$  by  $P(A \Delta B) = 0$  where  $A \Delta B$  denotes the symmetric difference of two sets.

- ★ Trick :  $(A_n \text{ i.o.}) = \left( \sum_{n=1}^{\infty} I_{A_n} = \infty \right)$

✓ Given  $\{A_n\}$  is independent, by setting  $\mathcal{F}_n = \sigma(A_1, \dots, A_n)$ , conditional Borel-Cantelli second lemma implies original Borel-Cantelli second lemma which is given by

$$\sum_n P(A_n) = \infty \Rightarrow P(A_n \text{ i.o.}) = 1$$

- Branching process (Galton-Watson process)

- Let  $\{\xi_i^n\}_{i \in \mathbb{N}, n \in \mathbb{N}}$  be i.i.d nonnegative integer-valued random variables.  
Define a Galton-Watson process  $\{Z_n\}_{n \in \mathbb{N} \cup \{0\}}$  as below :

$$Z_0 = 1$$

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

✓ The idea behind the definitions is that  $Z_n$  is the population in the  $n$ -th generation and each member of the  $n$ -th generation gives birth independently to an identically distributed number of offspring.

- $P(\xi_1^1 = k) \quad \forall k \in \mathbb{N} \cup \{0\}$  is called the offspring distribution.  $\mu = E(\xi_1^1)$  is the expected number of offspring per individual.

- Properties of the branching process

- Let  $\mathcal{F}_n = \sigma(\{\xi_i^m : i \in \mathbb{N}, 1 \leq m \leq n\}) \quad \forall n \in \mathbb{N}$ ,  $\mathcal{F}_0 = \{\phi, \Omega\}$ . If  $\mu = E(\xi_1^1) \in (0, \infty)$  then  $\{Z_n/\mu^n\}$  is a martingale w.r.t  $\{\mathcal{F}_n\}$  and  $E(Z_n) = \mu^n \quad \forall n \in \mathbb{N}$
- If  $\mu = E(\xi_1^1) \in (0, 1)$  then  $Z_n = 0$  for large enough  $n$ 's *a.s.* i.e. the species goes extinct.

- Inequalities for bounded stopping time

- If  $\{X_n\}$  is a submartingale and  $N$  is a stopping time with  $P(N \leq K) = 1$  for some  $K \in \mathbb{N}$  then

$$E(X_0) \leq E(X_N) \leq E(X_K)$$

✓ Since  $\{X_n\}$  is a submartingale,  $E(X_0) \leq E(X_j) \leq E(X_K)$  whenever  $0 \leq j \leq K$ . This thm tells us that similar inequality still holds true when the index is random.