

Lecture 6: Martingale

Tianpei Xie

Feb.2nd., 2023

Contents

1	Conditional Expectation	2
2	Martingale	4
2.1	Martingale, Sub-Martingale, Super-Martingale	4
2.2	Martingale Difference Sequence	5
2.3	Examples of Martingales	6
2.4	Doob's Decomposition	7
2.5	Stopping Time and Optional Sampling Theorem	8
2.6	Martingale Inequalities	9
2.7	Convergence Theorem	11

1 Conditional Expectation

- **Definition** (*Conditional Expectation*) [Resnick, 2013]

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $\mathcal{G} \subset \mathcal{F}$ be a sub- σ -algebra. Suppose $X \in L^1(\Omega, \mathcal{F}, \mathcal{P})$. There exists a function $\mathbb{E}[X|\mathcal{G}]$, called the **conditional expectation** of X **with respect to** \mathcal{G} such that

1. $\mathbb{E}[X|\mathcal{G}]$ is \mathcal{G} -**measurable** and **integrable with respect to** \mathcal{P} .
2. $\mathbb{E}[X|\mathcal{G}]$ satisfies **the functional equation**:

$$\int_G X d\mathcal{P} = \int_G \mathbb{E}[X|\mathcal{G}] d\mathcal{P}, \quad \forall G \in \mathcal{G}.$$

- **Remark** To *prove the existence* of such a random variable,

1. consider first the case of **nonnegative** X . Define a measure ν on \mathcal{G} by

$$\nu(G) = \int_G X d\mathcal{P} = \int_\Omega X \mathbf{1}_G d\mathcal{P}.$$

This measure is *finite* because X is *integrable*, and it is **absolutely continuous** with respect to \mathcal{P} . By the *Lebesgue-Radon-Nikodym Theorem*, there is a \mathcal{G} -measurable function f such that

$$\nu(G) = \int_G f d\mathcal{P}.$$

This f has properties (1) and (2).

2. If X is *not necessarily nonnegative*, $\mathbb{E}[X_+|\mathcal{G}] - \mathbb{E}[X_-|\mathcal{G}]$ clearly has the required properties.

- **Remark** As \mathcal{G} increases, condition (1) becomes **weaker** and condition (2) becomes **stronger**.

- **Remark** Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space, with $\mathcal{G} \subset \mathcal{F}$ a sub- σ -algebra, define

$$\mathcal{P}[A|\mathcal{G}] = \mathbb{E}[\mathbf{1}_A|\mathcal{G}]$$

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

- **Remark** By definition, the conditional expectation is a **Radon-Nikodym derivative** of $d\nu|_{\mathcal{G}} = X d\mathcal{P}|_{\mathcal{G}}$ w.r.t. $d\mathcal{P}|_{\mathcal{G}}$ within \mathcal{G} .

$$\mathbb{E}[X|\mathcal{G}] := \frac{X d\mathcal{P}|_{\mathcal{G}}}{d\mathcal{P}|_{\mathcal{G}}} = X|_{\mathcal{G}}.$$

Thus $\mathbb{E}[X|\mathcal{G}]$ is the **projection of X on sub σ -algebra \mathcal{G}** .

- **Remark** (*Conditioning on Random Variables*)

By definition, conditioning on random variables $(X_t, t \in T)$ on (Ω, \mathcal{B}) can be expressed as

$$\mathbb{E}[X|X_t, t \in T] \equiv \mathbb{E}[X|\sigma(X_t, t \in T)],$$

where $\sigma(X_t, t \in T)$ is the σ -algebra generated by the cylinder set

$$C_n[A] \equiv \{\omega : (X_t(\omega), 1 \leq t \leq n) \in A\} \in \mathcal{B}, \quad A \in \mathcal{B}(\mathbb{R}^n), \forall n$$

- **Remark** (*σ -Algebra Generated by Partition of Sample Space*)

As above, assume that the sub σ -algebra \mathcal{G} is generated by a **partition** B_1, B_2, \dots of Ω , then for $X \in L^1(\Omega, \mathcal{F}, \mathcal{P})$,

$$\mathbb{E}[X|B_i] = \int X d\mathcal{P}(X|B_i) = \int_{B_i} X d\mathcal{P} / \mathcal{P}(B_i)$$

where $\mathcal{P}(X|B_i)$ is the conditional probability defined in previous section. If $\mathcal{P}(B_i) = 0$, then $\mathbb{E}[X|B_i] = 0$. We claim that

1.

$$\mathbb{E}[X|\mathcal{G}] = \sum_{i=1}^{\infty} \mathbb{E}[X|B_i] \mathbb{1}_{B_i}, \quad a.s.$$

2. For any $A \in \mathcal{F}$,

$$\mathcal{P}(A|\mathcal{G}) = \sum_{i=1}^{\infty} \mathcal{P}(A|B_i) \mathbb{1}_{B_i}, \quad a.s.$$

- **Remark** Both $P[A|\mathcal{F}]$ and $\mathbb{E}[X|\mathcal{F}]$ are random variables from $\Omega \rightarrow \mathbb{R}$. Formally speaking,

$$\begin{aligned} P[(X, Y) \in A | \sigma(X)]_{\omega} &\equiv P[(X(\omega), Y) \in A] \\ &= P\{\omega' : (X(\omega), Y(\omega')) \in A\} \\ &\equiv f(X(\omega)) \\ &= \nu|_{\sigma(X)}(A) \\ \mathbb{E}[(X, Y) | \sigma(X)]_{\omega} &= \lim_{\substack{m(A) \rightarrow 0 \\ \omega \in A \in \sigma(X)}} \frac{P\{\omega' : (X(\omega), Y(\omega')) \in A\}}{m(A)} \end{aligned}$$

It is the expected value of X for someone who knows for each $E \in \mathcal{F}$, whether or not $\omega \in E$, which E itself remains unknown.

- **Proposition 1.1** (*Properties of Conditional Expectation*) [Resnick, 2013]

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $\mathcal{G} \subset \mathcal{F}$ be a sub- σ -algebra. Suppose $X, Y \in L^1(\Omega, \mathcal{F}, \mathcal{P})$ and $\alpha, \beta \in \mathbb{R}$.

1. (**Linearity**): $\mathbb{E}[\alpha X + \beta Y | \mathcal{G}] = \alpha \mathbb{E}[X | \mathcal{G}] + \beta \mathbb{E}[Y | \mathcal{G}]$;
2. (**Projection**): If X is \mathcal{G} -measurable, then $\mathbb{E}[X | \mathcal{G}] = X$ almost surely.
3. (**Conditioning on Indiscrete σ -Algebra**):

$$\mathbb{E}[X | \{\emptyset, \Omega\}] = \mathbb{E}[X].$$

4. (**Monotonicity**): If $X \geq 0$, then $\mathbb{E}[X | \mathcal{G}] \geq 0$ almost surely. Similarly, if $X \geq Y$, then $\mathbb{E}[X | \mathcal{G}] \geq \mathbb{E}[Y | \mathcal{G}]$ almost surely.
5. (**Modulus Inequality**):

$$|\mathbb{E}[X | \mathcal{G}]| \leq \mathbb{E}[|X| | \mathcal{G}].$$

6. (**Monotone Convergence Theorem**): If $\{X_n\}_{n=1}^\infty \subset L^1(\Omega, \mathcal{F}, \mathcal{P})$, $0 \leq X_1 \leq X_2 \leq \dots$ is a **monotone sequence of non-negative** random variables and $X_n \rightarrow X$ then

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{G}] = \mathbb{E}\left[\lim_{n \rightarrow \infty} X_n | \mathcal{G}\right] = \mathbb{E}[X | \mathcal{G}].$$

7. (**Fatou Lemma**): If $\{X_n\}_{n=1}^\infty \subset L^1(\Omega, \mathcal{F}, \mathcal{P})$, and $X_n \geq 0$ for all n , then

$$\mathbb{E}\left[\liminf_{n \rightarrow \infty} X_n | \mathcal{G}\right] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{G}]$$

8. (**Dominated Convergence Theorem**): If $\{X_n\}_{n=1}^\infty \subset L^1(\Omega, \mathcal{F}, \mathcal{P})$ and $|X_n| \leq Z$, where $Z \in L^1(\Omega, \mathcal{F}, \mathcal{P})$ is a random variable, $X_n \rightarrow X$ almost surely, then

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{G}] = \mathbb{E}\left[\lim_{n \rightarrow \infty} X_n | \mathcal{G}\right] = \mathbb{E}[X | \mathcal{G}], \quad a.s.$$

9. (**Product Rule**): If Y is \mathcal{G} -measurable,

$$\mathbb{E}[X Y | \mathcal{G}] = Y \mathbb{E}[X | \mathcal{G}], \quad a.s.$$

10. (**Smoothing**): For $\mathcal{F}_1 \subset \mathcal{F}_0 \subset \mathcal{F}$,

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X | \mathcal{F}_0] | \mathcal{F}_1] &= \mathbb{E}[X | \mathcal{F}_1] \\ \mathbb{E}[\mathbb{E}[X | \mathcal{F}_1] | \mathcal{F}_0] &= \mathbb{E}[X | \mathcal{F}_0]. \end{aligned}$$

Note that $\mathbb{E}[X | \mathcal{F}_1]$ is **smoother** than $\mathbb{E}[X | \mathcal{F}_0]$. Moreover

$$\mathbb{E}[X] = \mathbb{E}[X | \{\emptyset, \Omega\}] = \mathbb{E}[\mathbb{E}[X | \mathcal{F}_0] | \{\emptyset, \Omega\}] = \mathbb{E}[\mathbb{E}[X | \mathcal{F}_0]].$$

11. (**The Conditional Jensen's Inequality**). Let ϕ be a **convex** function, $\phi(X) \in L^1(\Omega, \mathcal{F}, \mathcal{P})$. Then almost surely

$$\phi(\mathbb{E}[X | \mathcal{G}]) \leq \mathbb{E}[\phi(X) | \mathcal{G}]$$

2 Martingale

2.1 Martingale, Sub-Martingale, Super-Martingale

- **Definition (Martingale)** [Resnick, 2013]

Let $\{X_n, n \geq 0\}$ be a stochastic process on (Ω, \mathcal{F}) and $\{\mathcal{F}_n, n \geq 0\}$ be a **filtration**; that is, $\{\mathcal{F}_n, n \geq 0\}$ is an *increasing sub σ -fields* of \mathcal{F}

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots \subseteq \mathcal{F}.$$

Then $\{(X_n, \mathcal{F}_n), n \geq 0\}$ is a **martingale (mg)** if

1. X_n is **adapted** in the sense that for each n , $X_n \in \mathcal{F}_n$; that is, X_n is \mathcal{F}_n -measurable.
2. $X_n \in L_1$; that is $\mathbb{E}[|X_n|] < \infty$ for $n \geq 0$.

3. For $0 \leq m < n$

$$\mathbb{E}[X_n \mid \mathcal{F}_m] = X_m, \quad \text{a.s.} \quad (1)$$

If the equality of (1) is replaced by \geq ; that is, things are getting better on the average:

$$\mathbb{E}[X_n \mid \mathcal{F}_m] \geq X_m, \quad \text{a.s.} \quad (2)$$

then $\{X_n\}$ is called a **sub-martingale (submg)** while if things are getting worse on the average

$$\mathbb{E}[X_n \mid \mathcal{F}_m] \leq X_m, \quad \text{a.s.} \quad (3)$$

$\{X_n\}$ is called a **super-martingale (supermg)**.

- **Remark** $\{X_n\}$ is ***martingale*** if it is *both* a ***sub*** and ***supermartingale***. $\{X_n\}$ is a ***super-martingale*** if and only if $\{-X_n\}$ is a ***submartingale***.
- **Remark** If $\{X_n\}$ is a ***martingale***, then $\mathbb{E}[X_n]$ is *constant*. In the case of a ***submartingale***, the mean increases and for a ***supermartingale***, the mean decreases.
- **Proposition 2.1** [Resnick, 2013]
If $\{(X_n, \mathcal{F}_n), n \geq 0\}$ is a ***(sub, super) martingale***, then

$$\{(X_n, \sigma(X_0, X_1, \dots, X_n)), n \geq 0\}$$

is also a ***(sub, super) martingale***.

2.2 Martingale Difference Sequence

- **Definition (Martingale Differences)**. [Resnick, 2013]
 $\{(d_j, \mathcal{B}_j), j \geq 0\}$ is a ***(sub, super) martingale difference sequence*** or a ***(sub, super) fair sequence*** if
 1. For $j \geq 0$, $\mathcal{B}_j \subset \mathcal{B}_{j+1}$.
 2. For $j \geq 0$, $d_j \in L_1$, $d_j \in \mathcal{B}_j$; that is, d_j is *absolutely integrable* and \mathcal{B}_j -measurable.
 3. For $j \geq 0$,

$$\begin{aligned} \mathbb{E}[d_{j+1} \mid \mathcal{B}_j] &= 0, & (\text{martingale difference / fair sequence}); \\ &\geq 0, & (\text{submartingale difference / subfair sequence}); \\ &\leq 0, & (\text{supermartingale difference / supfair sequence}) \end{aligned}$$

- **Theorem 2.2 (Construction of Martingale From Martingale Difference)**[Resnick, 2013]
If $\{(d_j, \mathcal{B}_j), j \geq 0\}$ is ***(sub, super) martingale difference sequence***, and

$$X_n = \sum_{j=0}^n d_j,$$

then $\{(X_n, \mathcal{B}_n), n \geq 0\}$ is a ***(sub, super) martingale***.

- **Theorem 2.3** (*Construction of Martingale Difference From Martingale*) [Resnick, 2013]

Suppose $\{(X_n, \mathcal{B}_n), n \geq 0\}$ is a (*sub, super*) *martingale*. Define

$$\begin{aligned} d_0 &:= X_0 - \mathbb{E}[X_0] \\ d_j &:= X_j - X_{j-1}, \quad j \geq 1. \end{aligned}$$

Then $\{(d_j, \mathcal{B}_j), j \geq 0\}$ is a (*sub, super*) *martingale difference sequence*.

- **Theorem 2.4** (*Orthogonality of Martingale Differences*). [Resnick, 2013]
If $\{(X_n, \mathcal{B}_n), n \geq 0\}$ is a *martingale* where X_n can be decomposed as

$$X_n = \sum_{j=0}^n d_j,$$

d_j is \mathcal{B}_j -measurable and $\mathbb{E}[d_j^2] < \infty$ for $j \geq 0$, then $\{d_j\}$ are *orthogonal*:

$$\mathbb{E}[d_i d_j] = 0 \quad i \neq j.$$

Proof: This is an easy verification: If $j > i$, then

$$\begin{aligned} \mathbb{E}[d_i d_j] &= \mathbb{E}[\mathbb{E}[d_i d_j | \mathcal{B}_i]] \\ &= \mathbb{E}[d_i \mathbb{E}[d_j | \mathcal{B}_i]] = 0. \quad \blacksquare \end{aligned}$$

A consequence is that

$$\mathbb{E}[X_n^2] = \mathbb{E}\left[\sum_{i=1}^n d_i^2\right] + 2 \sum_{0 \leq i < j \leq n} \mathbb{E}[d_i d_j] = \mathbb{E}\left[\sum_{i=1}^n d_i^2\right],$$

which is *non-decreasing*. From this, it seems likely (and turns out to be true) that $\{X_n^2\}$ is a *sub-martingale*.

2.3 Examples of Martingales

- **Example** (*Smoothing as Martingale*)

Suppose $X \in L_1$ and $\{\mathcal{B}_n, n \geq 0\}$ is an increasing family of sub σ -algebra of \mathcal{B} . Define for $n \geq 0$

$$X_n := \mathbb{E}[X | \mathcal{B}_n].$$

Then (X_n, \mathcal{B}_n) is a *martingale*. From this result, we see that $\{(d_n, \mathcal{B}_n), n \geq 0\}$ is a *martingale difference sequence* when

$$d_n := \mathbb{E}[X | \mathcal{B}_n] - \mathbb{E}[X | \mathcal{B}_{n-1}], \quad n \geq 1. \quad (4)$$

Proof: See that

$$\begin{aligned} \mathbb{E}[X_{n+1} | \mathcal{B}_n] &= \mathbb{E}[\mathbb{E}[X | \mathcal{B}_{n+1}] | \mathcal{B}_n] \\ &= \mathbb{E}[X | \mathcal{B}_n] \quad (\text{Smoothing property of conditional expectation}) \\ &= X_n \quad \blacksquare \end{aligned}$$

- **Example (*Sums of Independent Random Variables*)**

Suppose that $\{Z_n, n \geq 0\}$ is an *independent* sequence of integrable random variables satisfying for $n \geq 0$, $\mathbb{E}[Z_n] = 0$. Set

$$\begin{aligned} X_0 &:= 0, \\ X_n &:= \sum_{i=1}^n Z_i, \quad n \geq 1 \\ \mathcal{B}_n &:= \sigma(Z_0, \dots, Z_n). \end{aligned}$$

Then $\{(X_n, \mathcal{B}_n), n \geq 0\}$ is a *martingale* since $\{(Z_n, \mathcal{B}_n), n \geq 0\}$ is a *martingale difference sequence*.

- **Example (*Likelihood Ratios*)**

Suppose $\{Y_n, n \geq 0\}$ are *independent identically distributed* random variables and suppose the *true density* of Y_n is f_0 (The word “density” can be understood with respect to some fixed reference measure μ .) Let f_1 be *some other probability density*. For simplicity suppose $f_0(y) > 0$, for all y . For $n \geq 0$, define the likelihood ratio

$$\begin{aligned} X_n &:= \frac{\prod_{i=0}^n f_1(Y_i)}{\prod_{i=0}^n f_0(Y_i)} \\ \mathcal{B}_n &:= \sigma(Y_0, \dots, Y_n) \end{aligned}$$

Then (X_n, \mathcal{B}_n) is a *martingale*.

Proof: See that

$$\begin{aligned} \mathbb{E}[X_{n+1} | \mathcal{B}_n] &= \mathbb{E}\left[\left(\frac{\prod_{i=0}^n f_1(Y_i)}{\prod_{i=0}^n f_0(Y_i)}\right) \frac{f_1(Y_{n+1})}{f_0(Y_{n+1})} \mid Y_0, \dots, Y_n\right] \\ &= X_n \mathbb{E}\left[\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})} \mid Y_0, \dots, Y_n\right] \\ &= X_n \mathbb{E}\left[\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}\right] \quad (\text{by independence}) \\ &:= X_n \int \frac{f_1(y_{n+1})}{f_0(y_{n+1})} f_0(y_{n+1}) d\mu(y_{n+1}) = X_n. \quad \blacksquare \end{aligned}$$

- **Example (*Moment Generating Functions*)**

- **Example (*Method of Centering by Conditional Means*)**

- **Example (*Markov Chains*)**

2.4 Doob’s Decomposition

- **Definition (*Predictable and Increasing Process*)**

Given a process $\{X_n, n \geq 0\}$ and an increasing family of σ -algebras $\{\mathcal{B}_n, n \geq 0\}$. We call

$\{X_n, n \geq 0\}$ **predictable** if $X_0 \in \mathcal{B}_0 = \{\emptyset, \Omega\}$ and X_n is \mathcal{B}_{n-1} -**measurable** for each $n = 1, 2, \dots$

Call a process $\{A_n, n \geq 0\}$ an **increasing process** if $\{A_n\}$ is **predictable** and **almost surely**

$$0 = A_0 \leq A_1 \leq \dots \leq A_n \leq \dots$$

- **Theorem 2.5 (Doob Decomposition)** [Billingsley, 2008, Resnick, 2013]
Any **submartingale** $\{(X_n, \mathcal{B}_n), n \geq 0\}$ can be written in a **unique** way as the **sum** of a **martingale** $\{(M_n, \mathcal{B}_n), n \geq 0\}$ and an **increasing process** $\{A_n, n \geq 0\}$; that is

$$X_n = M_n + A_n, \quad n \geq 0, \quad a.s.$$

2.5 Stopping Time and Optional Sampling Theorem

- **Definition** Let $(X_t)_t$ be a stochastic process and let $\{\mathcal{B}_t, t \geq 0\}$ be an increasing family of σ -algebras. Denote $\mathcal{B} = \bigcup_{t \geq 0} \mathcal{B}_t$.

A random variable $T : (\Omega, \mathcal{B}) \rightarrow (\mathbb{N}_+ \cup \{+\infty\}, 2^{\mathbb{N}_+ \cup \{+\infty\}})$ is called a **stopping time** with respect to $\{\mathcal{B}_t, t \geq 0\}$, if $\mathbf{1}_{\{T=k\}}$ is \mathcal{B}_k -**measurable** for $\forall k \geq 0$; i.e.

$$\{T = k\} \in \mathcal{B}_k, \quad \forall k \geq 0.$$

- **Theorem 2.6 (Weak Version of Optional Sampling Theorem)** [Billingsley, 2008]
Let $\{(X_n, \mathcal{B}_n), n \geq 0\}$ be a **martingale** where $\mathcal{B}_n = \sigma(X_1, \dots, X_n)$ and let T be **stopping time** with respect to $\{\mathcal{B}_n, n \geq 0\}$. Suppose that **at least one** of the following conditions hold:

1. $T \leq n_0$, a.s. for some constant $n_0 > 0$;
2. $T < \infty$, a.s. and $|X_i| \leq K$ where $i \leq T$.

Then

$$\mathbb{E}[X_T] = \mathbb{E}[X_0].$$

- **Theorem 2.7 (Weak Version of Optional Sampling Theorem, again)** [Billingsley, 2008]
Let $\{(X_n, \mathcal{B}_n), n \geq 0\}$ be a **martingale** where $\mathcal{B}_n = \sigma(X_1, \dots, X_n)$ and let τ be **stopping time** with respect to $\{\mathcal{B}_n, n \geq 0\}$. Suppose $\mathbb{E}[\tau] < \infty$, and there exists $n \geq 0$ and some constant c such that

$$\mathbb{E}[|X_{n+1} - X_n| | \mathcal{B}_n] \leq c, \quad a.s. \text{ on } \{\tau \geq n\}.$$

Then $\mathbb{E}[|X_\tau|] < \infty$ and

$$\mathbb{E}[X_\tau] = \mathbb{E}[X_0].$$

- **Theorem 2.8 (Doob's Optional Sampling Theorem)** [Billingsley, 2008, Resnick, 2013]
Let $\{(X_n, \mathcal{B}_n), n \geq 0\}$ be a **martingale** and let S, T be $\{\mathcal{B}_n, n \geq 0\}$ -**stopping time** bounded by constant c , with $S \leq T \leq c$, a.s., then

$$\mathbb{E}[X_T | \mathcal{B}_S] = X_S, \quad a.s. \tag{5}$$

where

$$\mathcal{B}_S := \{A \in \mathcal{B} : A \cap \{S \leq n\} \in \mathcal{B}_n, \quad n = 0, 1, \dots\}$$

- **Theorem 2.9 (Wald's Identity)** [Billingsley, 2008, Resnick, 2013]

Consider random walk $S_0 = X_0$ and $S_n = X_0 + \xi_1 + \xi_2 + \dots + \xi_n$, $n = 1, 2, \dots$, where ξ_1, \dots, ξ_n are i.i.d. random variables such that $\mathbb{E}[|\xi_i|] < \infty$ and $\mathbb{E}[\xi_i] = \mu$. Let τ be a **stopping time** with respect to $\{\sigma(S_0, S_1, \dots, S_n), n \geq 0\}$ and $\mathbb{E}[\tau] < \infty$. Then

$$\mathbb{E}[S_\tau - S_0] = \mu \mathbb{E}[\tau]$$

- **Proposition 2.10 (Wald's Equation)** [Billingsley, 2008]

Let X_1, \dots, X_n be i.i.d. random variables with $\mathbb{E}[|X_i|] < \infty$ and $\mathbb{E}[X_i] = \mu < \infty$. Let τ be a **stopping time** with respect to $\{\sigma(X_1, \dots, X_n), n \geq 1\}$ satisfying $\mathbb{E}[\tau] < \infty$. Then

$$\mathbb{E}\left[\sum_{i=1}^{\tau} X_i\right] = \mathbb{E}[\tau] \mathbb{E}[X_i] \quad (6)$$

- **Proposition 2.11** [Resnick, 2013]

Let $\{(X_n, \mathcal{B}_n), n \geq 0\}$ be a **martingale** and let τ be a **stopping time** with respect to $\{\mathcal{B}_n, n \geq 0\}$. Then

$$\{X_{\tau \wedge t}, \quad t = 0, 1, \dots\}$$

is a **martingale** with respect to $\{\mathcal{B}_n, n \geq 0\}$, where $t \wedge \tau = \min\{t, \tau\}$.

2.6 Martingale Inequalities

- **Theorem 2.12 (Doob's Maximal Inequality)** [Billingsley, 2008]

Suppose $\{(X_n, \mathcal{B}_n), n \geq 0\}$ is a **sub-martingale**. Then for any $t > 0$,

$$\mathbb{P}\left\{\max_{k \leq n} X_k \geq t\right\} \leq \frac{1}{t} \mathbb{E}\left[X_n \mathbb{1}\left\{\max_{k \leq n} X_k \geq t\right\}\right] \quad (7)$$

- **Definition (Gambling Strategy)**

A **gambling strategy** with respect to an increasing family of σ -algebras $\{\mathcal{B}_n, n \geq 0\}$ is a sequence of random variables $\gamma_1, \gamma_2, \dots$ such that γ_n is \mathcal{B}_{n-1} -**measurable** for each $n = 1, 2, \dots$, where $\mathcal{B}_0 = \{\emptyset, \Omega\}$.

The random variable sequence $(\gamma_1, \gamma_2, \dots)$ is called **predictable**.

- **Definition (Upcrossing)**

Consider a sequence of random variables X_1, \dots, X_n, \dots such that X_n is \mathcal{B}_n -measurable for all n and two real numbers $a < b$. Define a gambling strategy $(\gamma_1, \gamma_2, \dots)$ by setting:

$$\begin{aligned} \gamma_0 &= 0; \\ \gamma_{n+1} &= \begin{cases} 1 & \text{if } \gamma_n = 0 \text{ and } X_n < 0 \\ 1 & \text{if } \gamma_n = 1 \text{ and } X_n \geq b \\ 0 & \text{otherwise} \end{cases} \quad \forall n = 1, 2, \dots \end{aligned}$$

Such strategy is called a **upcrossing strategy**.

- **Lemma 2.13** (*The Upcrossing Inequality Lemma*) [Billingsley, 2008]

Suppose $\{(X_n, \mathcal{B}_n), n \geq 0\}$ is a **super-martingale** and $a < b$. Then

$$\mathbb{E}[U_n[a, b]] \leq \frac{\mathbb{E}[(X_n - a)_+]}{b - a} = \frac{-\mathbb{E}[(X_n - a)_-]}{b - a} \quad (8)$$

where $U_n[a, b]$ denote **the total number of upcrossings** of $[a, b]$ up to time n ,

$$\begin{aligned} (X_n - a)_- &= \min\{X_n - a, 0\} = -\max\{a - X_n, 0\}; \\ (X_n - a)_+ &= \max\{X_n - a, 0\}. \end{aligned}$$

- **Theorem 2.14** (*Bernstein Inequality, Martingale Difference Sequence Version*)

[Wainwright, 2019]

Let $\{(D_k, \mathcal{B}_k), k \geq 1\}$ be a **martingale difference sequence**, and suppose that

$$\mathbb{E}[\exp(\lambda D_k) | \mathcal{B}_{k-1}] \leq \exp\left(\frac{\lambda^2 \nu_k^2}{2}\right)$$

almost surely for any $|\lambda| < 1/\alpha_k$. Then the following hold:

1. The sum $\sum_{k=1}^n D_k$ is **sub-exponential** with **parameters** $(\sqrt{\sum_{k=1}^n \nu_k^2}, \alpha_*)$ where $\alpha_* := \max_{k=1, \dots, n} \alpha_k$. That is, for any $|\lambda| < 1/\alpha_*$,

$$\mathbb{E}\left[\exp\left\{\lambda \left(\sum_{k=1}^n D_k\right)\right\}\right] \leq \exp\left(\frac{\lambda^2 \sum_{k=1}^n \nu_k^2}{2}\right)$$

2. The sum satisfies **the concentration inequality**

$$\mathbb{P}\left\{\left|\sum_{k=1}^n D_k\right| \geq t\right\} \leq \begin{cases} 2 \exp\left(-\frac{t^2}{2 \sum_{k=1}^n \nu_k^2}\right) & \text{if } 0 \leq t \leq \frac{\sum_{k=1}^n \nu_k^2}{\alpha_*} \\ 2 \exp\left(-\frac{t}{\alpha_*}\right) & \text{if } t > \frac{\sum_{k=1}^n \nu_k^2}{\alpha_*}. \end{cases} \quad (9)$$

- **Corollary 2.15** (*Azuma-Hoeffding Inequality, Martingale Difference*) [Wainwright, 2019]

Let $\{(D_k, \mathcal{B}_k), k \geq 1\}$ be a **martingale difference sequence** for which there are constants $\{(a_k, b_k)\}_{k=1}^n$ such that $D_k \in [a_k, b_k]$ almost surely for all $k = 1, \dots, n$. Then, for all $t \geq 0$,

$$\mathbb{P}\left\{\left|\sum_{k=1}^n D_k\right| \geq t\right\} \leq 2 \exp\left(-\frac{2t^2}{\sum_{k=1}^n (b_k - a_k)^2}\right) \quad (10)$$

- An important application of *Azuma-Hoeffding Inequality* concerns functions that satisfy a **bounded difference property**.

Definition (*Functions with Bounded Difference Property*)

Given vectors $x, x' \in \mathcal{X}^n$ and an index $k \in \{1, 2, \dots, n\}$, we define a new vector $x^{(-k)} \in \mathcal{X}^n$ via

$$x_j^{(-k)} = \begin{cases} x_j & j \neq k \\ x'_k & j = k \end{cases}$$

With this notation, we say that $f : \mathcal{X}^n \rightarrow \mathbb{R}$ satisfies ***the bounded difference inequality*** with parameters (L_1, \dots, L_n) if, for each index $k = 1, 2, \dots, n$,

$$\left| f(x) - f(x^{(-k)}) \right| \leq L_k, \quad \text{for all } x, x' \in \mathcal{X}^n. \quad (11)$$

- **Corollary 2.16** (*McDiarmid's Inequality / Bounded Differences Inequality*) [Wainwright, 2019]

Suppose that f satisfies ***the bounded difference property*** (11) with parameters (L_1, \dots, L_n) and that the random vector $X = (X_1, X_2, \dots, X_n)$ has ***independent*** components. Then

$$\mathbb{P} \{ |f(X) - \mathbb{E}[f(X)]| \geq t \} \leq 2 \exp \left(- \frac{2t^2}{\sum_{k=1}^n L_k^2} \right). \quad (12)$$

2.7 Convergence Theorem

- **Theorem 2.17** (*Doob's Martingale Convergence Theorem*) [Billingsley, 2008]

Suppose $\{(X_n, \mathcal{B}_n), n \geq 0\}$ is a ***super-martingale*** and suppose further that

$$\sup_n \mathbb{E} [|X_n|] < \infty.$$

Then there exists an integrable random variable X such that

$$\lim_{n \rightarrow \infty} X_n = X \quad a.s.$$

- **Remark** Note that a martingale is a super-martingale, so the above theorem is ***valid*** for ***martingale***. Also the above theorem is valid for ***sub-martingale*** when $\{(-X_n, \mathcal{B}_n), n \geq 0\}$ is considered.

References

Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.

Sidney I Resnick. *A probability path*. Springer Science & Business Media, 2013.

Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.