Lecture 6: Martingale

Tianpei Xie

Feb.2nd., 2023

Contents

1 Conditional Expectation

- **Definition** (***Conditional Expectation***) [Resnick, 2013]
Let $(\Omega, \mathscr{F}, \mathcal{P})$ be a probability space and $\mathscr{G} \subset \mathscr{F}$ be a sub-σ-algebra. Suppose $X \in L^1(\Omega, \mathscr{F}, \mathcal{P})$. There exists a function $\mathbb{E}\left[X|\mathscr{G}\right]$, called the ***conditional expectation*** of X ***with respect to*** $\mathscr{G}$ such that

  1. $\mathbb{E}\left[X|\mathscr{G}\right]$ is ***$\mathscr{G}$-measureable*** and ***integrable*** *with respect to* $\mathcal{P}$.
  2. $\mathbb{E}\left[X|\mathscr{G}\right]$ satisfies ***the functional equation***:
$$\int_G X d\mathcal{P} = \int_G \mathbb{E}\left[X|\mathscr{G}\right] d\mathcal{P}, \quad \forall\, G \in \mathscr{G}.$$

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

  1. consider first the case of ***nonnegative*** X. Define a measure ν on $\mathscr{G}$ by
$$\nu(G) = \int_G X d\mathcal{P} = \int_\Omega X \mathbb{1}_G \, d\mathcal{P}.$$
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 $\mathscr{G}$-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*, $\mathbb{E}\left[X_+|\mathscr{G}\right] - \mathbb{E}\left[X_-|\mathscr{G}\right]$ clearly has the required properties.

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

- **Remark** Let $(\Omega, \mathscr{F}, \mathcal{P})$ be a probability space, with $\mathscr{G} \subset \mathscr{F}$ a sub-σ-algebra, define
$$\mathcal{P}[A|\mathscr{G}] = \mathbb{E}\left[\mathbb{1}_A|\mathscr{G}\right]$$
for all $A \in \mathscr{F}$.

- **Remark** By definition, the conditional expectation is a ***Radon-Nikodym derivative*** of $d\nu|_{\mathscr{G}} = X d\mathcal{P}|_{\mathscr{G}}$ w.r.t. $d\mathcal{P}|_{\mathscr{G}}$ within $\mathscr{G}$.
$$\mathbb{E}\left[X|\mathscr{G}\right] := \frac{X d\mathcal{P}|_{\mathscr{G}}}{d\mathcal{P}|_{\mathscr{G}}} = X|_{\mathscr{G}}.$$
Thus $\mathbb{E}\left[X|\mathscr{G}\right]$ *is the* ***projection*** *of* X *on* ***sub*** ***σ-algebra*** $\mathscr{G}$.

- **Remark** (***Conditioning on Random Variables***)
By definition, conditioning on random variables $(X_t, t \in T)$ on $(\Omega, \mathscr{B})$ can be expressed as
$$\mathbb{E}\left[X|X_t, t \in T\right] \equiv \mathbb{E}\left[X|\sigma(X_t, t \in T)\right],$$
where $\sigma(X_t, t \in T)$ is the σ-algebra generated by the cylinder set
$$C_n[A] \equiv \{\omega : (X_t(\omega), 1 \le t \le n) \in A\} \in \mathscr{B}, \quad A \in \mathcal{B}(\mathbb{R}^n), \forall\, n$$

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

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

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

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

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

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

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

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

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

- **Proposition 1.1** ***(Properties of Conditional Expectation)*** *[Resnick, 2013]*
  Let $(\Omega, \mathscr{F}, \mathcal{P})$ be a probability space and $\mathscr{G} \subset \mathscr{F}$ be a sub-σ-algebra. Suppose $X, Y \in L^1(\Omega, \mathscr{F}, \mathcal{P})$ and $\alpha, \beta \in \mathbb{R}$.

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

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

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

$$|\,\mathbb{E}\left[X|\mathscr{G}\right]\,| \leq \mathbb{E}\left[\,|X|\,\,|\mathscr{G}\right].$$

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

$$\lim_{n\to\infty} \mathbb{E}\left[X_n|\mathscr{G}\right] = \mathbb{E}\left[\lim_{n\to\infty} X_n \middle| \mathscr{G}\right] = \mathbb{E}\left[X|\mathscr{G}\right].$$

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

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n \middle| \mathscr{G}\right] \leq \liminf_{n\to\infty} \mathbb{E}\left[X_n|\mathscr{G}\right]$$

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

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

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

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

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

$$\begin{aligned} \mathbb{E}\left[\mathbb{E}\left[X|\mathscr{F}_0\right]\,|\mathscr{F}_1\right] &= \mathbb{E}\left[X|\mathscr{F}_1\right] \\ \mathbb{E}\left[\mathbb{E}\left[X|\mathscr{F}_1\right]\,|\mathscr{F}_0\right] &= \mathbb{E}\left[X|\mathscr{F}_1\right]. \end{aligned}$$

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

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

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

$$\phi\left(\mathbb{E}\left[X|\mathscr{G}\right]\right) \leq \mathbb{E}\left[\phi(X)|\mathscr{G}\right]$$

2 Martingale

2.1 Martingale, Sub-Martingale, Super-Martingale

- **Definition** (***Martingale***) [Resnick, 2013]
  Let $\{X_n, n \geq 0\}$ be a stochastic process on $(\Omega, \mathscr{F})$ and $\{\mathscr{F}_n, n \geq 0\}$ be a ***filtration***; that is, $\{\mathscr{F}_n, n \geq 0\}$ is an *increasing sub σ-fields* of $\mathscr{F}$

$$\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \ldots \subseteq \mathscr{F}.$$

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

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

3. For $0 \le m < n$

$$\mathbb{E}\left[X_n \mid \mathscr{F}_m\right] = X_m, \quad \text{a.s.} \tag{1}$$

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

$$\mathbb{E}\left[X_n \mid \mathscr{F}_m\right] \ge X_m, \quad \text{a.s.} \tag{2}$$

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

$$\mathbb{E}\left[X_n \mid \mathscr{F}_m\right] \le X_m, \quad \text{a.s.} \tag{3}$$

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

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

  $$\{(X_n, \sigma\left(X_0, X_1, \ldots, X_n\right)), n \ge 0\}$$

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

2.2 Martingale Difference Sequence

- **Definition** (***Martingale Differences***). [Resnick, 2013]
  $\{(d_j, \mathscr{B}_j), j \ge 0\}$ is a ***(sub, super) martingale difference sequence*** or a ***(sub, super) fair sequence*** if

  1. For $j \ge 0$, $\mathscr{B}_j \subset \mathscr{B}_{j+1}$.
  2. For $j \ge 0$, $d_j \in L_1$, $d_j \in \mathscr{B}_j$; that is, d_j is *absolutely integrable* and $\mathscr{B}_j$*-measurable.*
  3. For $j \ge 0$,

  $$\begin{aligned} \mathbb{E}\left[d_{j+1} | \mathscr{B}_j\right] &= 0, && (\textit{martingale difference / fair sequence}); \\ &\ge 0, && (\textit{submartingale difference / subfair sequence}); \\ &\le 0, && (\textit{supmartingale difference / supfair sequence}) \end{aligned}$$

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

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

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

- **Theorem 2.3** *(**Construction of Martingale Difference From Martingale**) [Resnick, 2013]*
  *Suppose $\{(X_n, \mathscr{B}_n), n \geq 0\}$ 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 $\{(d_j, \mathscr{B}_j), j \geq 0\}$ is a **(sub, super) martingale difference sequence**.*

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

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

  *d_j is $\mathscr{B}_j$-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

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

  A consequence is that

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

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

2.3 Examples of Martingales

- **Example** (***Smoothing as Martingale***)
  Suppose $X \in L_1$ and $\{\mathscr{B}_n, n \geq 0\}$ is an increasing family of sub σ-algebra of $\mathscr{B}$. Define for $n \geq 0$

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

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

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

  Proof: See that

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

- **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

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

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

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

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

  Proof: See that

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

- **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 $\{\mathscr{B}_n, n \geq 0\}$. We call

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

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

$$0 = A_0 \leq A_1 \leq \ldots \leq A_n \leq \ldots.$$

- **Theorem 2.5** ***(Doob Decomposition)*** *[Billingsley, 2008, Resnick, 2013]*
Any ***submartingale*** $\{(X_n, \mathscr{B}_n), n \geq 0\}$ *can be written in a* ***unique*** *way as the* ***sum*** *of a* ***martingale*** $\{(M_n, \mathscr{B}_n), n \geq 0\}$ *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 Times and Optional Sampling Theorem

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

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

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

- **Theorem 2.6** ***(Weak Version of Optional Sampling Theorem)*** *[Billingsley, 2008]*
Let $\{(X_n, \mathscr{B}_n), n \geq 0\}$ *be a* ***martingale*** *where* $\mathscr{B}_n = \sigma(X_1, \ldots, X_n)$ *and let* T *be* ***stopping time*** *with respect to* $\{\mathscr{B}_n, n \geq 0\}$*. 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 $\{(X_n, \mathscr{B}_n), n \geq 0\}$ *be a* ***martingale*** *where* $\mathscr{B}_n = \sigma(X_1, \ldots, X_n)$ *and let* τ *be* ***stopping time*** *with respect to* $\{\mathscr{B}_n, n \geq 0\}$*. Suppose* $\mathbb{E}[\tau] < \infty$*, and there exists* $n \geq 0$ *and some constant* c *such that*

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

  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 $\{(X_n, \mathscr{B}_n), n \geq 0\}$ be a ***martingale*** *and let S, T be $\{\mathscr{B}_n, n \geq 0\}$-**stopping time** bounded by constant c, with $S \leq T \leq c$, a.s., then*

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

  where

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

- **Theorem 2.9** ***(Wald's Identity)*** *[Billingsley, 2008, Resnick, 2013]*
  *Consider random walk $S_n = X_0 + \xi_1 + \xi_2 + \ldots + \xi_n$, $n = 0, 1, 2, \ldots$, where $\xi_1, \ldots, \xi_n$ are i.i.d. random variables such that $\mathbb{E}\left[|\xi_i|\right] < \infty$ and $\mathbb{E}\left[\xi_i\right] = \mu$. Let τ be a $\{\sigma(S_0, S_1, \ldots, S_n), n \geq 0\}$-**stopping time** and $\mathbb{E}\left[\tau\right] < \infty$. Then*

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

- **Proposition 2.10** ***(Wald's Equation)*** *[Billingsley, 2008]*
  *Let $X_1, \ldots, X_n$ be i.i.d. random variables with $\mathbb{E}\left[|X_i|\right] < \infty$ and $\mathbb{E}\left[X_i\right] = \mu < \infty$. Let τ be a **stopping time** with respect to $\{\sigma\left(X_1, \ldots, X_n\right), n \geq 1\}$ satisfying $\mathbb{E}\left[\tau\right] < \infty$. Then*

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

- **Proposition 2.11** *) [Resnick, 2013]*
  *Let $\{(X_n, \mathscr{B}_n), n \geq 0\}$ be a **martingale** and let τ be a **stopping time** with respect to $\{\mathscr{B}_n, n \geq 0\}$. Then*

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

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

2.6 Martingale Inequalities

- **Theorem 2.12** ***(Doob's Maximal Inequality)*** *[Billingsley, 2008]*
  *Suppose $\{(X_n, \mathscr{B}_n), n \geq 0\}$ is a **sub-martingale**. Then for any $t > 0$,*

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

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

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

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

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

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

- **Lemma 2.13** *(**The Upcrossing Inequality Lemma**) [Billingsley, 2008]*
  *Suppose $\{(X_n, \mathscr{B}_n), n \geq 0\}$ is a **super-martingale** and $a < b$. Then*

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

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

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

- **Theorem 2.14** *(**Bernstein Inequality, Martingale Difference Sequence Version**) [Wainwright, 2019]*
  *Let $\{(D_k, \mathscr{B}_k), k \geq 1\}$ be a **martingale difference sequence**, and suppose that*

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

  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** $\left(\sqrt{\sum_{k=1}^n \nu_k^2}\ ,\ \alpha_*\right)$ where $\alpha_* := \max_{k=1,\ldots,n} \alpha_k$. 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***

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

- **Corollary 2.15** *(**Azuma-Hoeffding Inequality, Martingale Difference**)[Wainwright, 2019]*
  *Let $\{(D_k, \mathscr{B}_k), k \geq 1\}$ 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 $k = 1, \ldots, n$. Then, for all $t \geq 0$,*

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

- 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 $k \in \{1, 2, \ldots, n\}$, 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 $f : \mathcal{X}^n \to \mathbb{R}$ satisfies ***the bounded difference inequality*** with parameters $(L_1, \ldots, L_n)$ if, for each index $k = 1, 2, \ldots, n$,

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

- **Corollary 2.16** ***(McDiarmid's Inequality / Bounded Differences Inequality)****[Wainwright, 2019]*
  Suppose that f satisfies ***the bounded difference property*** (11) *with parameters* $(L_1, \ldots, L_n)$ *and that the random vector $X = (X_1, X_2, \ldots, X_n)$ has* ***independent*** *components. Then*

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

2.7 Convergence Theorem

- **Theorem 2.17** ***(Doob's Martingale Convergence Theorem)*** *[Billingsley, 2008]*
  Suppose $\{(X_n, \mathscr{B}_n), n \geq 0\}$ 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

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

- **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 $\{(-X_n, \mathscr{B}_n), n \geq 0\}$ 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.