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*** $B_1, B_2, \dots$ 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 $X \ge 0$, then $\mathbb{E}\left[X|\mathscr{G}\right] \ge 0$ almost surely. Similarly, if $X \ge Y$, then $\mathbb{E}\left[X|\mathscr{G}\right] \ge \mathbb{E}\left[Y|\mathscr{G}\right]$ almost surely.*
  5. *(**Modulus Inequality**):*
  $$\left|\,\mathbb{E}\left[X|\mathscr{G}\right]\,\right| \le \mathbb{E}\left[\,|X|\;|\mathscr{G}\right].$$

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

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

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

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

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

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

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*

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

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*

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

2 Martingale

2.1 Martingale, Sub-Martingale, Super-Martingale

- **Definition** (***Martingale***) [Resnick, 2013]
  Let $\{X_n, n \ge 0\}$ be a stochastic process on $(\Omega, \mathscr{F})$ and $\{\mathscr{F}_n, n \ge 0\}$ be a ***filtration***; that is, $\{\mathscr{F}_n, n \ge 0\}$ 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 $\{(X_n, \mathscr{F}_n), n \ge 0\}$ 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 $n \ge 0$.

3. For $0 \leq m < n$

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

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

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

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

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

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

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

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

2.2 Martingale Difference Sequence

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

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

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

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

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

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

- **Proposition 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*

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

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

- **Proposition 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**:*

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

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

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

  A consequence is that

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

  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***.

- **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 $n \geq 0$

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

  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

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

  Proof: See that

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

- **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

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

  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

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

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

  Proof: See that

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

2.3 Martingale Inequalities

- **Proposition 2.5** ***(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*

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

  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*** $\left(\sqrt{\sum_{k=1}^{n} \nu_k^2}\ ,\ \alpha_*\right)$ *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***

$$\mathbb{P}\left\{\left|\sum_{k=1}^{n} D_k\right| \geq t\right\} \leq \begin{cases} 2\exp\left(-\frac{t^2}{2\sum_{k=1}^{n}\nu_k^2}\right) & \textit{if } 0 \leq t \leq \frac{\sum_{k=1}^{n}\nu_k^2}{\alpha_*} \\ 2\exp\left(-\frac{t}{\alpha_*}\right) & \textit{if } t > \frac{\sum_{k=1}^{n}\nu_k^2}{\alpha_*}. \end{cases} \tag{5}$$

- **Corollary 2.6** *(**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 $t \geq 0$,*

$$\mathbb{P}\left\{\left|\sum_{k=1}^{n} D_k\right| \geq t\right\} \leq 2\exp\left(-\frac{2t^2}{\sum_{k=1}^{n}(b_k - a_k)^2}\right) \tag{6}$$

- 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 $k = 1, 2, \ldots, n$,

$$\left|f(x) - f(x^{(-k)})\right| \leq L_k, \quad \text{for all } x, x' \in \mathcal{X}^n. \tag{7}$$

- **Corollary 2.7** *(**McDiarmid's Inequality / Bounded Differences Inequality**)[Wainwright, 2019]*
  Suppose that f satisfies ***the bounded difference property*** (7) *with parameters $(L_1, \ldots, L_n)$ and that the random vector $X = (X_1, X_2, \ldots, X_n)$ has* ***independent*** *components. Then*

$$\mathbb{P}\left\{|f(X) - \mathbb{E}\left[f(X)\right]| \geq t\right\} \leq 2\exp\left(-\frac{2t^2}{\sum_{k=1}^{n} L_k^2}\right). \tag{8}$$

References

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.