Lecture 12

Instructor: Quan Zhou

For more details about the materials covered in this note, see Chapter 10.4 of Resnick [3] and Chapter 4.2 of Durrett [2].

12.1 Martingales

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space.

Definition 12.1. $\{\mathcal{F}_i\}_{i\geq 1}$ is called a filtration, if it is an non-decreasing sequence of σ-algebras, i.e. each $\mathcal{F}_i$ is a σ-algebra on Ω and $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}$. We often assume the index i starts from 0 and let $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Definition 12.2. A sequence of random variables $\{X_i\}_{i\geq 1}$ is said to be adapted to $\{\mathcal{F}_i\}_{i\geq 1}$ if $X_i \in \mathcal{F}_i$ for each i.

Definition 12.3. A sequence of random variables $\{X_i\}_{i\geq 1}$ adapted to $\{\mathcal{F}_i\}_{i\geq 1}$ is said to be a martingale w.r.t. $\{\mathcal{F}_i\}_{i\geq 1}$, if for each i, we have (i) $E|X_i| < \infty$; (ii) $E[X_{i+1} \mid \mathcal{F}_i] = X_i$.

Example 12.1. Consider tossing a coin infinitely many times. The sample space is $\Omega = \{\omega = (\omega_1, \omega_2, \dots) : \ \omega_i \in \{\mathrm{H}, \mathrm{T}\}\}$. We can define a filtration $\{\mathcal{F}_n\}_{n\geq 1}$ by letting $\mathcal{F}_n$ be the σ-algebra "generated by" the first n tosses; i.e., $\mathcal{F}_n$ represents the information we have after observing n tosses. By convention, we often let $\mathcal{F}_0 = \{\emptyset, \Omega\}$ to represent no information. To simplify the notation, let A_H denote the set of all sequences beginning with H, i.e.

$$A_H = \{\omega = (\omega_1, \omega_2, \dots) : \ \omega_1 = \mathrm{H}\}.$$

$A_T, A_{HH}, A_{HT}, \dots$ are defined similarly. Then, we have

$$\begin{aligned}
\mathcal{F}_1 &= \{\emptyset, \Omega, A_H, A_T\}, \\
\mathcal{F}_2 &= \{\emptyset, \Omega, A_H, A_T, A_{HH}, A_{HT}, A_{TH}, A_{TT}, \\
&\quad A_{HH} \cup A_{TH}, A_{HH} \cup A_{TT}, A_{HT} \cup A_{TH}, A_{HT} \cup A_{TT}, A^c_{HH}, A^c_{HT}, A^c_{TH}, A^c_{TT}\}.
\end{aligned}$$

Example 12.2. Let $Z_1, Z_2, \dots$ be a sequence of i.i.d. random variables with $E[Z_i] = 0$ (and thus $E|Z_i| < \infty$). Let $S_n = \sum_{i=1}^n Z_i$ and define $\mathcal{F}_n = \sigma(Z_1, \dots, Z_n)$. Then $\{S_n\}_{n\geq 1}$ is a martingale.

12.2 Martingale concentration inequalities

Lemma 12.1. *For any distribution F on $[0,1]$ with mean p and any convex function $\varphi : [0,1] \to \mathbb{R}$, we have*

$$\int_0^1 \varphi(x) F(dx) \le p\varphi(1) + (1-p)\varphi(0).$$

That is, if X is a random variable with distribution F, then $E[\varphi(X)]$ is maximized when F is the Bernoulli distribution with mean p.

Proof. Let X be a random variable with distribution F, and let U be a uniform random variable on $[0,1]$ independent of X. Define another random variable $Y = \mathbb{1}_{\{U \le X\}}$, which follows a Bernoulli distribution. Further,

$$E[Y] = E[E[Y \mid X]] = E[E[\mathbb{1}_{\{U \le X\}} \mid X]] = E[X] = p.$$

(Conditioning on $X = x$, U is still uniformly distributed.) Of course you can also directly compute $E[Y]$. Since U, X are independent, their joint distribution is given by the product measure and then by Fubini's Theorem,

$$E[Y] = \int_{[0,1]} \int_{[0,1]} \mathbb{1}_{\{u \le x\}} \mathsf{P}_X(dx) \mathsf{P}_U(du) = E[X],$$

where P_X, P_U denote the distribution of X and U respectively. Since Y is bounded, by Jensen's inequality (the conditional version),

$$E[\varphi(Y) \mid X] \ge \varphi(E[Y \mid X]) = \varphi(X).$$

Taking expectation on both sides, we obtain the asserted inequality. □

Theorem 12.1 (Azuma's inequality)**.** *Let $X_0, X_1, \dots$ be a martingale sequence (X_0 can be understood as the initial value of this process) such that $|X_i - X_{i-1}| \le c_i < \infty$, a.s. for $i = 1, 2, \dots$. Then,*

$$\mathsf{P}(X_n - X_0 \ge t) \le \exp\left(-\frac{t^2}{2\sum_{i=1}^n c_i^2} \right).$$

Remark 12.1. Let $Z_i = X_i - X_{i-1}$ for $i = 1, \dots, n$. Then, $X_n - X_0 = \sum_{i=1}^n Z_i$. If $Z_1, \dots, Z_n$ are i.i.d. with mean zero, we can also apply Hoeffding's inequality to obtain the above result.

Proof. First, apply the Chernoff bound to obtain

$$\mathsf{P}(X_n - X_0 \geq t) \leq e^{-\lambda t} E[e^{\lambda(X_n - X_0)}], \qquad \forall\, \lambda > 0.$$

The increments $|X_i - X_{i-1}|$ $(i = 1, 2, \dots)$ are not independent. But we can use conditioning.

$$\begin{aligned} E[e^{\lambda(X_n - X_0)}] &= E[E[e^{\lambda(X_n - X_0)} \mid \mathcal{F}_{n-1}]] \\ &= E[e^{\lambda(X_{n-1} - X_0)}\, E[e^{\lambda(X_n - X_{n-1})} \mid \mathcal{F}_{n-1}]]. \end{aligned}$$

By assumption $X_n - X_{n-1}$ is bounded on $(-c_n, c_n)$ and $E[X_n - X_{n-1} \mid \mathcal{F}_{n-1}] = 0$ since $\{X_n\}$ is a martingale. Hence, the conditional distribution of $\Delta_n = (X_n - X_{n-1} + c_n)/(2c_n)$ is a distribution on $[0, 1]$ with mean $1/2$.

Therefore, by Lemma 12.1 (recall that exp is convex),

$$E[e^{\lambda \Delta_n} \mid \mathcal{F}_{n-1}] \leq \frac{1}{2}\left(e^{\lambda} + 1\right).$$

Since $X_n - X_{n-1} = 2c_n\Delta_n - c_n$,

$$E[e^{\lambda(X_n - X_{n-1})} \mid \mathcal{F}_{n-1}] \leq \frac{1}{2} e^{-\lambda c_n}\left(e^{2\lambda c_n} + 1\right) = \cosh(\lambda c_n) \leq e^{\lambda^2 c_n^2/2}.$$

The last inequality can be proved by comparing the corresponding Taylor series expansions. Applying the above inequality iteratively, we get

$$e^{-\lambda t} E[e^{\lambda(X_n - X_0)}] \leq \exp\left(\frac{\lambda^2}{2}\sum_{i=1}^{n} c_i^2 - \lambda t\right).$$

Azuma's inequality is then proved by letting $\lambda = t/(\sum c_i^2)$. □

Theorem 12.2 (Efron-Stein inequality)**.** *Let $X_1, \dots, X_n$ be n independent random variables and let $Y = f(X_1, \dots, X_n)$ for some Borel function f such that $E|Y| < \infty$. Define $E_i[Y] = E[Y \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n]$. Then,*

$$\mathrm{Var}(Y) \leq \sum_{i=1}^{n} E\left(\{Y - E_i[Y]\}^2\right).$$

Proof. Define a filtration by letting $\mathcal{F}_i = \sigma(X_1, \dots, X_i)$. Define $\Delta_i = E[Y \mid \mathcal{F}_i] - E[Y \mid \mathcal{F}_{i-1}]$. By the properties of conditional expectation,

$$E[\Delta_i \mid \mathcal{F}_{i-1}] = 0,$$

and thus $\{E[Y \mid \mathcal{F}_i]\}_{i\geq 1}$ is a martingale w.r.t. $\{\mathcal{F}_i\}_{i\geq 1}$. Further,

$$\begin{aligned}\sum_{i=1}^n \Delta_i &= \sum_{i=1}^n (E[Y \mid \mathcal{F}_i] - E[Y \mid \mathcal{F}_{i-1}]) \\ &= E[Y \mid \mathcal{F}_n] - E[Y \mid \mathcal{F}_0] = Y - E[Y].\end{aligned}$$

Next, we claim that

$$\mathrm{Var}[Y] = E\left\{(Y - E[Y])^2\right\} = E\left[\left(\sum_{i=1}^n \Delta_i\right)^2\right] = \sum_{i=1}^n E[\Delta_i^2].$$

This is because for any $i > j$, we have

$$E[\Delta_i \Delta_j] = E[E[\Delta_i \Delta_j \mid \mathcal{F}_j]] = E[\Delta_j \, E[\Delta_i \mid \mathcal{F}_j]] = 0.$$

We claim that we have

$$E[Y \mid \mathcal{F}_{i-1}] = E[E[Y \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n] \mid \mathcal{F}_i]. \tag{1}$$

Note that conditioning on $\mathcal{F}_i$ just means to condition on $X_1, \dots, X_i$. To prove (1), let $U = (X_1, \dots, X_{i-1})$ and $V = E[Y \mid U, X_{i+1}, \dots, X_n]$. Then,

$$E[E[Y \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n] \mid \mathcal{F}_i] = E[V \mid \sigma(U, X_i)].$$

Observing that (V, U) is $\sigma(U, X_{i+1}, \dots, X_n)$-measurable, by Theorem 7.2, we know that (V, U) is independent of X_i. By Proposition 9.2 (iii),[1] we find that $E[V \mid \sigma(U, X_i)] = E[V \mid U]$, which is equal to $E[Y \mid \mathcal{F}_{i-1}]$ by the tower property of conditional expectation.

By Jensen's inequality for conditional expectation,

$$\begin{aligned}\Delta_i^2 &= (E[Y \mid \mathcal{F}_i] - E[Y \mid \mathcal{F}_{i-1}])^2 \\ &= (E\{Y - E[Y \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n] \mid \mathcal{F}_i\})^2 \\ &\leq E\left(\{Y - E[Y \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n]\}^2 \mid \mathcal{F}_i\right).\end{aligned}$$

Taking expectation on both sides we get

$$E[\Delta_i^2] \leq E\left(\{Y - E[Y \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n]\}^2\right),$$

which completes the proof. □

[1] In Proposition 9.2 (iii), we only considered the case of three random variables. But the extension to more than three random variables is straightforward.

Example 12.3. Consider the simple random walk $S_n = X_1 + X_2 + \cdots + X_n$ where $\{X_i\}_{i \geq 1}$ is an i.i.d. sequence of random variables and $\mathsf{P}(X_i = 1) = \mathsf{P}(X_i = -1) = 1/2$. Then $\{S_n\}_{n \geq 1}$ is a martingale with bounded increments. Hence, by Azuma's inequality,

$$\mathsf{P}(S_n > t) \leq e^{-t^2/2n} \qquad \text{i.e.} \quad \mathsf{P}(S_n > \sqrt{2n \log n}) \leq \frac{1}{n}.$$

This can also be obtained from Hoeffding's inequality.

References

[1] S Boucheron, G Lugosi, and P Massart. Concentration inequalities: A nonasymptotic theory of independence, 2013.

[2] Rick Durrett. *Probability: Theory and Examples*, volume 49. Cambridge university press, 2019.

[3] Sidney Resnick. *A Probability Path*. Springer, 2019.