6 Martingale, Concentration and Stopping Times

6.1 Martingale

The notion of martingale is used to describe fair games. Consider a gambler who wins X_i dollars in the i-th round of a sequence of bets. If in each round, the game is fair, then $\mathbb{E}[X_i] = 0$ regardless of the history. The variables $\{X_i\}$ are not necessarily independent, but if we use Z_i to denote the amount of money he won after i-th round, then clearly for every i, it holds that $\mathbb{E}[Z_i \mid Z_1, \dots, Z_{i-1}] = Z_{i-1}$. Of course, since the sequence $X_1, \dots, X_i$ contains same information (about the probability space) as $Z_1, \dots, Z_i$, the above property is equivalent to $\mathbb{E}[Z_i \mid X_1, \dots, X_{i-1}] = Z_{i-1}$. This is exactly the definition of *martingale*. For convenience, from now on we will use $\overline{X}_{0,n} = (X_0, \dots, X_n)$ to simplify our notations.

Definition 6.1. Martingale
A (discrete) stochastic process $\{X_n\}_{n\geq 0}$ is a *martingale* if

$$\forall n \geq 0, \qquad \mathbb{E}[X_{n+1} \mid X_0, \dots, X_n] = X_n .$$

We say $\{Z_n\}_{n\geq 0}$ is a martingale with respect to another sequence $\{X_n\}_{n\geq 0}$ if

$$\forall n \geq 0, \qquad \mathbb{E}[Z_{n+1} \mid \overline{X}_{0,n}] = Z_n .$$

More formally, given a sequence of random variables $\{X_n\}_{n\geq 0}$, let $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$ be the minimum σ-algebra generated by $\overline{X}_{0,n}$. Then $\{\mathcal{F}_n\}_{n\geq 0}$ satisfies

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \cdots .$$

In particular, a sequence of σ-algebra $\{\mathcal{F}_n\}_{n\geq 0}$ satisfying the above condition is called a *filtration*. We say $\{Z_n\}_{n\geq 0}$ is a *martingale* with respect to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$ if

1. $\forall n \geq 0$, Z_n is $\mathcal{F}_n$-measurable;
2. $\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] = Z_n$.

Simlarly, we say $\{Z_n\}_{n\geq 0}$ is a *supermartingale* if

$$\forall n \geq 0, \qquad \mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] \leq Z_n ,$$

and a *submartingale* if

$$\forall n \geq 0, \qquad \mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] \geq Z_n .$$

Example 6.1.
Consider a one-dimensional random walk on $\mathbb{Z}$ starting from 0. The probability to left and the probability to right are both $1/2$ at each step. Denote by a uniform random variable $X_n \in \{-1,+1\}$ the n-th step. Let $Z_n = \sum_{i=1}^{n} X_n$. Then $\{Z_n\}$ is a martingale with respect to $\{X_n\}$.
It is easy to verify that

1. Z_n is measurable given the σ-algebra $\sigma(\overline{X}_{0,n})$;
2. $\mathbb{E}[Z_{n+1} \mid \overline{X}_{0,n}] = \mathbb{E}[Z_n + X_{n+1} \mid \overline{X}_{0,n}] = Z_n + \mathbb{E}[X_{n+1} \mid \overline{X}_{0,n}] = Z_n$.

Example 6.2.
Pólya's Urn
Suppose there are some white balls and black balls in an urn. All of these balls are identical except the colors. Consider the following stochastic process: each round we pick a ball uniformly at random and observe its color; then we return the ball, and add an additional ball of the same color into the urn. We repeat the process, and our goal is to study the sequence of colors of the selected balls.
W.l.o.g. assume that we start from only one white ball and one black ball in the urn, and the index of rounds starts from 3. Then after round n, there are exactly n balls in the urn. Let X_n be the number of white balls after round n, and $Z_n = X_n/n$ is the ratio of white balls after round n. We claim that Z_n is a martingale.
Clearly $Z_2 = 1/2$ and $\mathbb{E}[Z_n] = 1/2$ since white balls and black balls are symmetric. We now compute $\mathbb{E}[Z_{n+1} \mid \overline{X}_{2,n}]$. Note that at round $n+1$, we pick a white ball with probability Z_n. Thus,

$$\begin{aligned}
\mathbb{E}[Z_{n+1} \mid \overline{X}_{2,n}] &= \frac{1}{n+1} \cdot \mathbb{E}[X_{n+1} \mid \overline{X}_{2,n}] \\
&= \frac{1}{n+1} \cdot (Z_n \cdot (X_n + 1) + (1 - Z_n) \cdot X_n) \\
&= \frac{1}{n+1} \cdot (X_n + Z_n) \ = Z_n \,.
\end{aligned}$$

If $\{Z_n\}_{n\geq 0}$ is a martingale w.r.t. $\{X_n\}_{n\geq 0}$, then the following property is immediate.

Proposition 6.1.
For any $n \geq 1$, $\mathbb{E}[Z_n] = \mathbb{E}[Z_0]$.

Proof. Note that by the definition,

$$\mathbb{E}[Z_n] = \mathbb{E}[\mathbb{E}[Z_n \mid \overline{X}_{0,n-1}]] = \mathbb{E}[Z_{n-1}].$$

Then the proposition follows by induction on n. □

6.2 Concentration Inequalities

Since we've known the expectation of Z_n, we now consider the concentration of Z_n.

Theorem 6.2. Azuma-Hoeffding Inequality
Let $Z_0, Z_1, \ldots, Z_n$ be a martingale such that $Z_i - Z_{i-1} \in [a_i, b_i]$ for any $i \in [n]$. Then,

$$\mathbf{Pr}[Z_n - Z_0 \geq \lambda] \leq \exp\left(-\frac{2\lambda^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

More generally, if Z_i conditioned on $Z_0, \ldots, Z_{i-1}$ lies inside an interval of length c_i (the interval may depends on $Z_0, \ldots, Z_{i-1}$, but its length is unpper bounded), then

$$\Pr[Z_n - Z_0 \geq \lambda] \leq \mathrm{e}^{-2\lambda^2/(c_1^2+\ldots+c_n^2)}.$$

Let $X_i = Z_i - Z_{i-1}$. It is equivalent to $X_i \in [a_i, b_i]$, $\mathbb{E}[X_i] = 0$, and $\{Z_n\}$ is a martingale w.r.t. $\{X_n\}$.

Remark 6.1.
Applying Auzma's inequality to $-Z_0$ and $-Z_n$, it gives

$$\Pr[|Z_n - Z_0| \geq \lambda] \leq 2\mathrm{e}^{-2\lambda^2/(c_1^2+\ldots+c_n^2)}.$$

Now we sketch a proof of the Azuma-Hoeffding, which is quite similar to our proof of the Hoeffding inequality. Recall when we were trying to prove the Hoeffding inequality, the most difficult part is to estimate the moment generating function, namely the term

$$\mathbb{E}\left[\mathrm{e}^{\lambda Z_n}\right] = \mathbb{E}\left[\prod_{i=1}^n \mathrm{e}^{\lambda X_i}\right].$$

There we applied the independent properties of random variables and obtain

$$\mathbb{E}\left[\prod_{i=1}^n \mathrm{e}^{\lambda X_i}\right] = \prod_{i=1}^n \mathbb{E}\left[\mathrm{e}^{\lambda X_i}\right].$$

Then by Hoeffding's lemma

$$\mathbb{E}\left[\mathrm{e}^{\lambda X_i}\right] \leq \mathrm{e}^{-\frac{\lambda(b_i-a_i)^2}{8}}.$$

In the case of Azuma-Hoeffding, we can use the property of martingales instead of independence to obtain a similar bound. To see this, we have

$$\begin{aligned}\mathbb{E}\left[\prod_{i=1}^n \mathrm{e}^{\lambda X_i}\right] &= \mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^n \mathrm{e}^{\lambda X_i} | \overline{X}_{0,n-1}\right]\right] \\ &= \mathbb{E}\left[\prod_{i=1}^{n-1} \mathrm{e}^{\lambda X_i} \mathbb{E}\left[\mathrm{e}^{\lambda X_n} \mid \overline{X}_{0,n-1}\right]\right].\end{aligned}$$

The bounds then follows by an induction argument and a conditional expectation version of Hoeffding lemma:

$$\mathbb{E}\left[\mathrm{e}^{\lambda X_n} \mid \overline{X}_{0,n-1}\right] \leq \mathrm{e}^{-\lambda^2(b_i-a_i)^2/8}.$$

The proof is almost the same as our proof of Hoeffding lemma in the last lecture.

An important family of martingale is the *Doob Sequence*.

Definition 6.2. Doob Sequence
Let $X_1, \ldots, X_n$ be a sequence of (unnecessarily independent) random variables and $f(\overline{X}_{1,n}) = f(X_1, \ldots, X_n) \in \mathbb{R}$ be a function. For $i \geq 0$, we define

$$Z_i = \mathbb{E}[f(\overline{X}_{1,n}) \mid \overline{X}_{1,i}].$$

The sequence $\{Z_i\}_{i=0,\ldots,n}$ is called a **Doob sequence** or **Doob martingale**.

We can see that

$$Z_0 = \mathbb{E}[f(\overline{X}_{1,n})];$$
$$Z_n = f(\overline{X}_{1,n}).$$

Therefore, Z_n is the value of the function given the input $\overline{X}_n$ and Z_0 is the average of the function value without any knowledge about the input. The sequence $\{Z_i\}_{i\geq 0}$ can be viewed as our estimation of the function value provided more and more information as i increases.

Proposition 6.3.
A Doob sequence $\{Z_i\}_{0\leq i\leq n}$ is a martingale w.r.t. $\overline{X}_{1,n}$.

Proof.

$$\mathbb{E}[Z_i \mid \overline{X}_{1,i-1}] = \mathbb{E}[\mathbb{E}[f(\overline{X}_{1,n}) \mid \overline{X}_{1,i}] \mid \overline{X}_{1,i-1}] = \mathbb{E}[f(\overline{X}_{1,n}) \mid \overline{X}_{1,i-1}] = Z_{i-1}. \quad \square$$

We now consider the following example. Suppose there are g green balls and r red balls in a bin and we want to estimate the ratio $\frac{r}{r+g}$ by drawing balls. There are two scenarios.

- Draw balls with replacement. Let $X_i = \mathbf{1}[\text{the } i\text{-th ball is red}]$. Let $X = \sum_{i=1}^n X_i$. Then clearly each $X_i \sim \text{Ber}\left(\frac{r}{r+g}\right)$ and $\mathbb{E}X = n \cdot \frac{r}{r+g}$. Since all X_is are independent, we can directly apply Hoeffding's inequality to obtain

$$\Pr[|X - \mathbb{E}X| \geq t] \leq 2\exp\left(-\frac{2t^2}{n}\right).$$

- Draw balls without replacement. Again we let $Y_i = \mathbf{1}[\text{the } i\text{-th ball is red}]$, then unlike the case of drawing with replacement, variables in $\{Y_i\}$ are dependent. Let $Y = \sum_{i=1}^n Y_i$. We first calculate $\mathbb{E}Y$.
For every $i \geq 1$, $\mathbb{E}Y_i$ is the probability that the i-th draw is a red ball. Note that drawing without replacement is equivalent to first drawing a uniform permutation of $r+g$ balls and drawing each ball one by one in that order. Therefore, the probability of $Y_i = 1$ is $\frac{r\cdot(r+g-1)!}{(r+g)!} = \frac{r}{r+g}$. So we have $\mathbb{E}Y = n \cdot \frac{r}{r+g}$.
Next, we consider the concentration of Y. We apply Azuma-Hoeffding for a certain martingale. Consider the n-ary function $f(y_1, y_2, \ldots, y_n) = \sum_{i=1}^n y_i$ and the Doob sequence of f. That is, let $Z_i = \mathbb{E}[f(\overline{Y}_{1,n}) \mid \overline{Y}_{1,i}]$, then we know $\{Z_i\}_{0\leq i\leq n}$ is a martingale. In order to fit the condition of Azuma-Hoeffding, note that

$$Z_i = (Z_i - Z_{i-1}) + (Z_{i-1} - Z_{i-2}) + \cdots + (Z_1 - Z_0) + Z_0$$

We let $X_i \triangleq Z_i - Z_{i-1}$ for $1 \leq i \leq n$, then

$$Z_n - Z_0 = Z_n - \mathbb{E}f = \sum_{i=1}^n X_i.$$

We need to bound the width of $|X_i| = |Z_i - Z_{i-1}|$. By definition,

$$Z_i - Z_{i-1} = \mathbb{E}[f(\overline{Y}_{1,n}) \mid \overline{Y}_{1,i}] - \mathbb{E}[f(\overline{Y}_{1,n}) \mid \overline{Y}_{1,i-1}].$$

If we use S_i to denote the number of 1's among $\overline{Y_i}$, namely $S_i = \sum_{j=1}^{i} Y_j$, then

$$\mathbb{E}[f(\overline{Y}_{1,n}) \mid \overline{Y}_{1,i}] = \mathbb{E}[f(\overline{Y}_{1,n}) \mid S_i] = S_i + (n-i) \cdot \frac{r - S_i}{g + r - i}.$$

Therefore $S_i = S_{i-1} + Y_i$ and

$$\begin{aligned} Z_i - Z_{i-1} &= \left(S_i + (n-i) \cdot \frac{r - S_i}{g+r-i}\right) - \left(S_{i-1} + (n-i+1) \cdot \frac{r - S_{i-1}}{g+r-i+1}\right) \\ &= \frac{g+r-n}{g+r-i}\left(Y_i + \frac{S_{i-1} - r}{g+r-i+1}\right). \end{aligned}$$

Note that $r \geq S_{i-1}$ and $g \geq (i-1) - S_{i-1}$, we have

$$\begin{aligned} Z_i - Z_{i-1} &\leq \frac{g+r-n}{g+r-i}\left(1 + \frac{S_{i-1} - r}{g+r-i+1}\right) \leq \frac{g+r-n}{g+r-i} \leq 1, \\ Z_i - Z_{i-1} &\geq \frac{g+r-n}{g+r-i}\left(\frac{S_{i-1} - r}{g+r-i+1}\right) \geq -\frac{g+r-n}{g+r-i} \geq -1. \end{aligned}$$

Therefore $-1 \leq X_i \leq 1$ and we can apply Azuma-Hoeffding to $Z_n - Z_0$ to obtain

$$\Pr\left[|Y - \mathbb{E}Y| \geq t\right] \leq 2\exp\left(-\frac{t^2}{2n}\right).$$

The Doob sequence we used in the Balls-into-Bags example is a very powerful and general tool to obtain concentration bounds. For a model defined by n random variables $X_1, \ldots, X_n$ and any quantity $f(X_1, \ldots, X_n)$ that we want to estimate, we can apply the Azuma-Hoeffding inequality to the Doob sequence of f. As shown in the previous example, the quality of the bound relies on the *width* of the martingale.

Let us first repeat the argument in the previous example. The Doob sequence is $Z_i = \mathbb{E}[f(\overline{X}_{0,n}) \mid \overline{X}_{0,i}]$ for every $0 \leq i \leq n$. For every $0 \leq i \leq n$, we let

$$S_i = Z_i - Z_0 = (Z_1 - Z_0) + \cdots + (Z_i - Z_{i-1}) = X_1 + \cdots + X_i,$$

where $X_j = Z_j - Z_{j-1}$. Then we apply Azuma-Hoeffding to $S_n = Z_n - Z_0 = f(\overline{X}_{0,n}) - \mathbb{E}[f(\overline{X}_{0,n})]$.

We need to determine the width of each X_i. This is relatively easy if the function f and the variables $\{X_i\}_{1 \leq i \leq n}$ enjoy certain nice properties.

Definition 6.3. Lipschitz function

A function $f(x_1, \cdots, x_n)$ satisfies c-Lipschitz condition if $\forall x_1, \ldots x_n$, $\forall i \in [n]$ and y_i,

$$|f(x_1, \cdots, x_i, \cdots, x_n) - f(x_1, \cdots, y_i, \cdots, x_n)| \leq c.$$

The McDiarmid's inequality is the application of Azuma-Hoeffding inequality to Lipschitz f and indepdenent $\{X_i\}$.

a.k.a. bounded-difference inequality

Theorem 6.4. McDiarmid's inequality

Let f be a function on n variables satisfying c-Lipschitz condition and $X_1, \cdots, X_n$ be n independent variables. Then we have

$$\Pr\left[\left|f(X_1, \cdots, X_n) - \mathbb{E}[f(X_1, \cdots, X_n)]\right| \geq t\right] \leq 2\mathrm{e}^{-\frac{2t^2}{nc^2}}.$$

Proof. We use f and $\{X_i\}_{i\geq 1}$ to define a Doob martingale $\{Z_i\}_{i\geq 1}$:

$$\forall i: Z_i = \mathbb{E}[f(\overline{X}_n) \mid \overline{X}_i].$$

Let

$$X_i = Z_i - Z_{i-1} = \mathbb{E}[f(\overline{X}) \mid \overline{X}_i] - \mathbb{E}[f(\overline{X}) \mid \overline{X}_{i-1}].$$

Next we try to determine the width of $Z_i - Z_{i-1}$. Clearly

$$Z_i - Z_{i-1} \geq \inf_x \left\{ \mathbb{E}[f(\overline{X}) \mid \overline{X}_{i-1}, X_i = x] - \mathbb{E}[f(\overline{X}) \mid \overline{X}_{i-1}] \right\},$$

and

$$Z_i - Z_{i-1} \leq \sup_y \left\{ \mathbb{E}[f(\overline{X}) \mid \overline{X}_{i-1}, X_i = y] - \mathbb{E}[f(\overline{X}) \mid \overline{X}_{i-1}] \right\}.$$

The gap between the upper bound and the lower bound is

$$\sup_{x,y} \left\{ \mathbb{E}[f(\overline{X}) \mid \overline{X}_{i-1}, X_i = y] - \mathbb{E}[f(\overline{X}) \mid \overline{X}_{i-1}, X_i = x] \right\}.$$

For every x, y and $\sigma_1, \ldots, \sigma_n$, we have

$$|f(\sigma_1, \ldots, \sigma_{i-1}, x, \sigma_{i+1}, \ldots, \sigma_n) - f(\sigma_1, \ldots, \sigma_{i-1}, y, \sigma_{i+1}, \ldots, \sigma_n)| \leq c.$$

Using the independence of $X_1, \ldots, X_n$, it follows that

$$\begin{aligned}
&\mathbb{E}[f(\overline{X}) \mid \overline{X}_{i-1}, X_i = y] - \mathbb{E}[f(\overline{X}) \mid \overline{X}_{i-1}, X_i = x] \\
&= \sum_{\sigma_{i+1}, \ldots, \sigma_n} \Big(\Pr\Big[\wedge_{i+1 \leq j \leq n} X_j = \sigma_j \mid \wedge_{1 \leq j \leq i-1} X_j = \sigma_j, X_i = y\Big] \cdot f(\sigma_1, \ldots, \sigma_{i-1}, y, \sigma_{i+1}, \ldots, \sigma_n) \\
&\quad - \Pr\Big[\wedge_{i+1 \leq j \leq n} X_j = \sigma_j \mid \wedge_{1 \leq j \leq i-1} X_j = \sigma_j, X_i = x\Big] \cdot f(\sigma_1, \ldots, \sigma_{i-1}, x, \sigma_{i+1}, \ldots, \sigma_n) \Big) \\
&= \sum_{\sigma_{i+1}, \ldots, \sigma_n} \Pr\Big[\wedge_{i+1 \leq j \leq n} X_j = \sigma_j\Big] \cdot (f(\sigma_1, \ldots, \sigma_{i-1}, y, \sigma_{i+1}, \ldots, \sigma_n) - f(\sigma_1, \ldots, \sigma_{i-1}, x, \sigma_{i+1}, \ldots, \sigma_n)) \\
&\leq c.
\end{aligned}$$

Applying Azuma-Hoeffding, we have

$$\Pr[|Z_n - Z_0| \geq t] = \Pr\Big[|f(X_1, \cdots, X_n) - \mathbb{E}[f(X_1, \cdots, X_n)]| \geq t\Big] \leq 2e^{-\frac{2t^2}{nc^2}}. \qquad \square$$

Let us examine two applications of McDiarmid's inequality.

Example 6.3.
Pattern Matching
Let $B \in \{0,1\}^k$ be a fixed string. For a random string $X \in \{0,1\}^n$, what is the expected number of occurrences of B in X?
The expectation of X can be easily calculated using the linearity of expectations. We define n independent random variables $X_1, \cdots, X_n$, where X_i denotes i-th character of X. Let $f(X_1, \cdots, X_n)$ be the number of occurrences of B in X. Note that there are at most $n-k+1$ occurrences of B in X, and we can enumerate the first position of each occurrence. By the linearity of expectation, we have

$$\mathbb{E}f = \frac{n-k+1}{2^k}.$$

We can then use McDarmid's inequality to show that f is well-concentrated. To see this, we note that variables in $\{X_i\}$ are independent and the function f is k-Lipschitz: If we change one bit of X, the number of occurrences changes at most k. Therefore

$$\Pr[|Z_n - Z_0| \geq t] = \Pr[|f - \mathbb{E}f| \geq t] \leq 2e^{-\frac{2t^2}{nk^2}}.$$

In random graphs, we have two classical martingales:

- Edge-exposure martingale: $\mathbb{E}\left[f(\mathcal{G}(n,p))|X_0, X_1, \ldots, X_{\binom{n}{2}}\right]$, where each variable symbolizes an edge;
- Vertex-exposure martingale: $\mathbb{E}[f(\mathcal{G}(n,p))|X_0, X_1, \ldots, X_n]$, where each variable symbolizes a vertex.

There is a trade-off between the length and the difference bound.

Example 6.4.
Chromatic Number of $\mathcal{G}(n,p)$
Another application of McDiarmid's Inequality is to establish the concentration of chromatic number for Erdős-Rényi random graphs $\mathcal{G}(n,p)$. Recall the notation $\mathcal{G}(n,p)$ specifies a distribution over all undirected simple graphs with n vertices. In the model, each of the $\binom{n}{2}$ possible edges exists with probability p independently.
For a graph $G \sim \mathcal{G}(n,p)$, we use $\chi(G)$ to denote its chromatic number, the minimum number q so that G can be properly colored using q colors. There are different ways to represent G using random variables.

- The most natural way is to introduce a variable X_e for every pair of vertices $e = \{u,v\} \in E$ where $X_e = \mathbf{1}[\text{the edge } e \text{ exists in } G]$. Then $\{X_e\}$ are independent and the chromatic number can be written as a function $\chi\left(X_{e_1}, X_{e_2}, \ldots, X_{e_{\binom{n}{2}}}\right)$. It is easy to see that χ is 1-Lipschitz as removing to adding one edge can only change the chromatic number by at most one. So by McDarmid's inequality, we have

$$\Pr[|\chi - \mathbb{E}[\chi]| \geq t] \leq 2\mathrm{e}^{-2t^2/\binom{n}{2}}.$$

  However, this bound is not satisfactory as we need to set $t = \Theta(n)$ in order to upper bound the RHS by a constant.

- We can encode the graph G in a more efficient way while reserving the Lipschitz and the independence property. Suppose the vertex set of G is $\{v_1, \ldots, v_n\}$. We define n random variables $Y_1, \cdots, Y_n$, where Y_i encodes the edges between v_i and $\{v_1, \cdots, v_{i-1}\}$. Once $Y_1, \cdots, Y_n$ are given, the graph is known, so the chromatic number can be written as a function $\chi(Y_1, \ldots, Y_n)$. Since Y_i only involves the connections between v_i and $v_1, \cdots, v_{i-1}$, the n variables are independent. It is also easy to see that if X_i changes, the chromatic number changes at most one. Hence χ is 1-Lipschitz as well. Applying McDiarmid's inequality we have

$$\Pr[|\chi - \mathbb{E}[\chi]| \geq t] \leq 2\mathrm{e}^{-2t^2/n}.$$

6.3 Stopping Time and Optional Stopping

Suppose $\{Z_n\}$ is a martingale. We've already known that $\forall n \geq 0$, $\mathbb{E}[Z_n] = \mathbb{E}[Z_0]$. However, if τ is a random variable, could we conclude that $\mathbb{E}[Z_\tau] = \mathbb{E}[Z_0]$?

Unfortunately the answer is "no" in general! For example, consider a one-dimensional random walk starting from 0. Let Z_n be the position after the n-th step, and τ be the first time that $Z_\tau = 1$. It is clear that $\mathbb{E}[Z_\tau] = 1 \neq Z_0$. Another example is to define τ

as $\arg\max_{1\le t\le 100}|Z_t|$, the time to reach the furthest position in the first 100 steps. Obviously we have $\mathbb{E}[Z_\tau] > 0 = Z_0$.

To determine under which condition we could conclude $\mathbb{E}[Z_\tau] = \mathbb{E}[Z_0]$, let's formalize *stopping time* first.

Definition 6.4. Stopping time
Let $\tau \in \mathbb{N} \cup \{\infty\}$ is a random variables. We say τ is a *stopping time* defined on a filtration $\{\mathcal{F}_n\}_{n\ge 0}$ if

$$\forall n \in \mathbb{N}, \qquad \mathbb{1}_{[\tau\le n]} \text{ is } \mathcal{F}_n\text{-measurable}\,.$$

In other words, for every n the event $\{\tau \le n\}$ is in $\mathcal{F}_n$.

Now consider the filtration generated by a stochastic process $\{X_n\}_{n\ge 0}$. We say τ is a stopping time if the proposition $[\tau \le n]$ is determined by $\overline{X}_{0,n}$ for all n.

Theorem 6.5. Optional Stopping Theorem
Suppose that $\{X_n\}$ is a martingale with respect to a filtration $\{\mathcal{F}_n\}$ and τ is a stopping time with respect to the same filtration. Then $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$ if at leat one of the following holds

1. τ is bounded;
2. $\Pr[\tau < \infty] = 1$ and $\exists M$ such that $|X_n| \le M$ for all $n < \tau$;
3. $\mathbb{E}[\tau] < \infty$ and $\exists c$ such that $\mathbb{E}[|X_{n+1} - X_n| \mid \mathcal{F}_n] \le c$ for all $n < \tau$.

The proof of the optional stopping theorem is left as an exercise. We now introduce some examples and applications of this theorem.

Example 6.5.
Suppose that in a villiage, every family keeps having children until they give birth to a boy. If we further assume that the natural birth sex ratio is uniform and every family only gives birth to a child at a time, what is the birth sex ratio in this villiage?
Fix a family. Let $X_n \in \{-1, +1\}$ denote whether the n-th child is a boy, and $Z_n = \sum_{i=1}^{n} X_i$ denote the number of boys more than girls. Then we define τ by $\tau = \min\{n : X_n = 1\}$. Clearly $\{Z_n\}$ is a martingale with respect to $\{X_n\}$ and τ is a stopping time.
Note that τ is the time of success in a Bernoulli trial and has a geometric distribution. Thus $\mathbb{E}[\tau] < \infty$. Combining with the fact that $|Z_{n+1} - Z_n| = |X_{n+1}| = 1$, it justifies Condition 3 in the Optional Stopping Theorem. So we conclude that $\mathbb{E}[Z_\tau] = \mathbb{E}[Z_0] = 0$ and hence the birth sex ratio in this villiage is still $1:1$.
Suppose that their strategy has changed. Every family keeps giving birth to children until their sons are more than their daughters. Then the optional stopping theorem cannot be applied any longer, since now the stopping time τ has infinite expectation and Z_n is unbounded. To see this, note that τ is the hitting time of 1 in a one-dimensional random walk starting from 0, that is, the hitting time of a *null recurrent* state.

Example 6.6.
One-Dimensional Random Walk with Two Absorbing Barriers
Consider a one-dimensional random walk (starting from 0) with two absorbing barriers $-a$ and b. There are two natural questions:

1. What is the probability of stopping at $-a$ (or b)?
2. What is the expected number of steps before stopping?

Let $X_n \in \{-1,+1\}$ be a uniform random variable, $Z_{n+1} = Z_n + X_n$, and $\tau = \min\{n\colon Z_n = -a \vee Z_n = b\}$. Then $\{Z_n\}$ is a martingale w.r.t. $\{X_n\}$, and τ is a stopping time.
Note that $|Z_n|$ is bounded. So, in order to apply the optional stopping theorem, we should prove that $\Pr[\tau < \infty] = 1$.
Since the probability of ending within the next $a+b$ steps is at least $2^{-(a+b)}$ no matter where the current position is, we claim that the random walk ends in finite steps with probability 1. It follows that $\mathbb{E}[Z_\tau] = \mathbb{E}[Z_0] = 0$. That is

$$-a \cdot \Pr[\text{ending at } -a] + b \cdot \Pr[\text{ending at } b] = 0\,,$$

which yields that the probability of ending at $-a$ and the probability of ending at b are $b/(a+b)$ and $a/(a+b)$, respectively.
We also define $\{Y_n\}_{n\geq 0}$ (which is a common trick) by

$$Y_n \triangleq Z_n^2 - n\,.$$

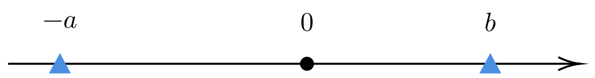
We claim that $\{Y_n\}$ is a martingale w.r.t. $\{X_n\}$. Now we can use Condition 3 in the Optional Stopping Theorem. It implies that $\mathbb{E}[Y_\tau] = \mathbb{E}[Y_0] = 0$. By the linearity of expectation, $\mathbb{E}[Y_\tau] = \mathbb{E}[Z_\tau^2] - \mathbb{E}[\tau]$. It follows that

$$\begin{aligned}\mathbb{E}[\tau] = \mathbb{E}[Z_\tau^2] &= a^2 \cdot \Pr[\text{ending at } -a] + b^2 \cdot \Pr[\text{ending at } b] \\ &= a^2 \cdot \frac{b}{a+b} + b^2 \cdot \frac{a}{a+b} = ab\,.\end{aligned}$$

Finally, we prove our claim, namely, $\{Y_n\}$ is a martingale with respect to $\{X_n\}$. By definition, we obtain that

$$\begin{aligned}\mathbb{E}[Y_{n+1} \mid \overline{X}_{0,n}] &= \mathbb{E}[Z_{n+1}^2 - (n+1) \mid \overline{X}_{0,n}] \\ &= \mathbb{E}[(Z_n + X_{n+1})^2 - (n+1) \mid \overline{X}_{0,n}] \\ &= Z_n^2 - n - 1 + 2Z_n \cdot \mathbb{E}[X_{n+1} \mid \overline{X}_{0,n}] + \mathbb{E}[X_{n+1}^2 \mid \overline{X}_{0,n}] \\ &= Z_n^2 - n = Y_n\,,\end{aligned}$$

which completes our proof.



Another example is the pattern matching problem. Suppose that there is a $\{\mathsf{H},\mathsf{T}\}$-string P of length ℓ. We flip a coin consecutively until the last ℓ results form exactly the same string as P. How many times do we flip the coin?

Note that if we flip the coin N times and observe the string consisting of N results. No matter which pattern we choose, the expected number of occurrence (i.e., expected number of substrings exactly the same as P of the resulting string) is $(N-$

$\ell+1)/2^\ell$ (by the linearity of expectation). However, if we would like to compute the first time that pattern P occurs, the pattern itself has an impact on the expected time.

Intuitively, let's consider two patterns HT and HH. Assume that the first flipping result is H. Then we consider what happens if the second result fails. Suppose that the desired pattern is HT and H appears. Altough we fail, we obtain an H. However, if the desired pattern is HH and the second flipping result is T, then we obtain nothing and the first two flips are a waste. So we should believe that the expected times of the first occurrence of these two patterns are different.

Now we can use the optional stopping theorem to solve this problem. Let $P = p_1p_2\dots p_\ell$. For every $n \geq 0$, assume that before $(n+1)$-th flipping there is a new gambler G_{n+1} coming with 1 unit of money to bet that the following ℓ result (i.e., the $(n+1)$-th to $(n+\ell)$-th results) are exactly the same as P. At the $(n+k)$-th flipping, G_{n+1} will bet that the result is p_k by an "all in" strategy, that is, if the $(n+k)$-th result is p_k then G_{n+1} will have twice as much money as before; otherwise they will lose all. Suppose that the patter $P=$ HTHTH and the flipping results are HTHHTHTH. The following table shows the total money of each gambler after flipping. Let X_t be the

Gambler	H	T	H	H	T	H	T	H	Money	
1	H	T	H	T					0	$1\to2\to4\to8\to0$
2		H							0	$1\to0$
3			H	T					0	$1\to2\to0$
4				H	T	H	T	H	32	$1\to2\to4\to8\to16\to32$
5					H				0	$1\to0$
6						H	T	H	8	$1\to2\to4\to8$
7							H		0	$1\to0$
8								H	2	$1\to2$

result of t-th flipping, $M_i(t)$ denote the money that G_i has after t-th flipping, and

$$Z_t = \sum_{i=1}^{t} M_i(t) - 1$$

be the total *net income* of all gamblers after t-th flipping. It is easy to verify that $\{M_i(t)\}_{t\geq0}$ is a martingale with respect to $\{X_n\}$ since

$$\mathbb{E}[M_i(t+1) \mid \overline{X}_{0,n}] = \frac{1}{2}\cdot(2M_i(t)) + \frac{1}{2}\cdot 0 = M_i(t)\,.$$

Then by the linearity of expectation we conclude that $\{Z_n\}$ is a martingale with respect to the flipping results $\{X_n\}$ since $\mathbb{E}[M_i(t)] = 0$. Let τ be the stopping time defined by the first time that some gambler wins, namely, the first time that P occurs in the flipping results. Applying Condition 2 of the Optional Stopping Theorem we obtain that $\mathbb{E}[Z_\tau] = \mathbb{E}[Z_0] = 0$.

We complete our solution by pointing out that G_i lose all for all $i \leq \tau-\ell$ and $M_i(\tau) = 2^{\tau-i+1}\cdot\chi_{\tau-i+1}$ for all $\tau-\ell+1 \leq i \leq \tau$, where χ_j is defined by

$$\chi_j = \mathbb{1}_{[p_1\cdots p_j = p_{\ell-j+1}\cdots p_\ell]}\,.$$

Hence,

$$0 = \mathbb{E}[Z_\tau] = \sum_{i=1}^{\tau}\mathbb{E}[M_i(\tau)] - \mathbb{E}[\tau] = \sum_{i=\tau-\ell+1}^{\tau} M_i(\tau) - \mathbb{E}[\tau] = \sum_{i=1}^{\ell}\chi_i\cdot 2^i - \mathbb{E}[\tau]\,.$$

6.4 Wald's Equation

In practice, we often need to analyze the (expected) running time of following procedure where both cond and compute() are random.

```
while cond do
    compute();
end while
```

Assume the i-th calling of compute() costs X_i time and the algorithm terminates after T iterations. Then the total running time is $N \triangleq \sum_{i=1}^{T} X_i$. Suppose X_is are independently and identically distributed as a random variable X. The Wald's equation gives a formulate for $\mathbb{E}N$.

Theorem 6.6. Wald's Equation
Suppose $X_1, X_2, \ldots$ are non-negative, independent, identically distributed random variables with the same distribution as X, T is a stopping time for $X_1, X_2, \ldots$, and both $\mathbb{E}[T], \mathbb{E}[X] < \infty$. Then we have

$$\mathbb{E}\left[\sum_{i=1}^{T} X_i\right] = \mathbb{E}[T] \cdot \mathbb{E}[X].$$

Proof. For $i \geq 1$, let $Z_i := \sum_{j=1}^{i} (X_j - \mathbb{E}X)$. Clearly the sequence $Z_1, Z_2, \ldots$ is a martingale with respect to $X_1, X_2, \ldots$ and $\mathbb{E}Z_i = 0$. Then we have

$$\begin{aligned}\mathbb{E}[|Z_{i+1} - Z_i| \mid \mathcal{F}_i] &= \mathbb{E}[|X_{i+1} - \mathbb{E}X| \mid \mathcal{F}_i] \\ &\leq \mathbb{E}[X_{i+1} + \mathbb{E}X \mid \mathcal{F}_i] \\ &\leq 2\mathbb{E}[X].\end{aligned}$$

We know that $\mathbb{E}T, \mathbb{E}X < \infty$, and therefore OST applies: $\mathbb{E}[Z_T] = \mathbb{E}[Z_1] = 0$.

Then it follows that

$$\begin{aligned}\mathbb{E}[Z_t] = \mathbb{E}\left[\sum_{j=1}^{T} (X_j - \mathbb{E}X)\right] &= \mathbb{E}\left[\sum_{i=1}^{T} X_i - T\mathbb{E}X\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{T} X_i\right] - \mathbb{E}[T] \cdot \mathbb{E}[X] = 0.\end{aligned}$$

□

Let us consider an application of Wald's equation. There are n senders and one receiver. In each round, each sender sends a packet to the receiver with probability $\frac{1}{n}$. Since all senders share the same channel, if there are multiple packets sent at the same time, all of them will fail. The question is, on average, how many rounds are required so that each sender can successfully send at least one packet.

We let X_i be the variable indicating how long the receiver needs to get another packet after he has received $i-1$ ones (counting packets from repeated sender). And let T be the number of packets received when first time the receiver receives at least one packet from each sender. The quantity we are interested in is

$$N \triangleq \sum_{i=1}^{T} X_i.$$

Clearly $X_1, X_2, \ldots$ are independently and identically distributed, and $\mathbb{E}T$ is finite. Therefore $\mathbb{E}N = \mathbb{E}T \cdot \mathbb{E}X$ by Wald's equation.

Note that by the definition, T is the number of coupons in the coupon collector's problem we met before. So $\mathbb{E}T = nH_n = \Theta(n \log n)$.

On the otherhand, $X_i \sim \mathtt{Geom}(p)$ with

$$p = n \cdot \frac{1}{n}\left(1 - \frac{1}{n}\right)^{n-1} \approx \mathrm{e}^{-1}$$

Therefore,

$$\mathbb{E}[N] = \mathbb{E}[T] \cdot \mathbb{E}[X_i] \approx \mathrm{e}nH_n.$$