Class 13-14: Martingales
Online Class

- Attend synchronous class (if possible), participate, ask questions

On Zoom:
- Use the largest display you have. Your phone is not a good choice
- Join Zoom with video ON and audio Muted
- View options: Fit to Window
- To participate:
  - Use Raise Hand in Reactions
  - Unmute and talk
  - Chat to Everyone or Host.
Hoeffding’s Bound

Theorem

Let $X_1, \ldots, X_n$ be independent random variables with $\mathbb{E}[X_i] = \mu_i$ and $\Pr(B_i \leq X_i \leq B_i + c_i) = 1$, then

$$\Pr(\left| \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i \right| \geq \epsilon) \leq e^{-\frac{2\epsilon^2}{\sum_{i=1}^{n} c_i^2}}$$

Do we need independence?
Martingales

Definition

A sequence of random variables \( Z_0, Z_1, \ldots \) is a **martingale** with respect to the sequence \( X_0, X_1, \ldots \) if for all \( n \geq 0 \) the following hold:

1. \( Z_n \) is a function of \( X_0, X_1, \ldots, X_n \);
2. \( E[|Z_n|] < \infty \);
3. \( E[Z_{n+1}|X_0, X_1, \ldots, X_n] = Z_n \);

Definition

A sequence of random variables \( Z_0, Z_1, \ldots \) is a **martingale** when it is a martingale with respect to itself, that is

1. \( E[|Z_n|] < \infty \);
2. \( E[Z_{n+1}|Z_0, Z_1, \ldots, Z_n] = Z_n \);
Definition

\[ E[Y \mid Z = z] = \sum_y y \Pr(Y = y \mid Z = z) , \]

where the summation is over all \( y \) in the range of \( Y \).

Note that \( E[Y \mid Z] \) is a random variable (a function of \( Z \)).

Lemma

For any random variables \( X \) and \( Y \),

\[ E[X] = E_Y[E_X[X \mid Y]] = \sum_y \Pr(Y = y)E[X \mid Y = y] , \]

where the sum is over all values in the range of \( Y \).
Lemma

For any random variables $X$ and $Y$, 

$$E[X] = E_Y[E_X[X \mid Y]] = \sum_y \Pr(Y = y)E[X \mid Y = y],$$

where the sum is over all values in the range of $Y$.

Proof.

\[
\begin{align*}
\sum_y \Pr(Y = y)E[X \mid Y = y] &= \sum_y \Pr(Y = y) \sum_x x \Pr(X = x \mid Y = y) \\
&= \sum_x \sum_y x \Pr(X = x \mid Y = y) \Pr(Y = y) \\
&= \sum_x \sum_y x \Pr(X = x \cap Y = y) = \sum_x x \Pr(X = x) = E[X].
\end{align*}
\]
Example

\[ Y \] - the number of students attending class, \( Y \sim B(n, p) \)

\[ X \] - the number of questions asked in class is \( X \mid Y = y \sim B(\lfloor \sqrt{y} \rfloor, q) \).

All events are independent

\[ E[X \mid Y = y] = q \lfloor \sqrt{y} \rfloor \] - a constant

\[ E[X \mid Y] = q \lfloor \sqrt{Y} \rfloor \] - a random variable

\[
E[X] = E_Y[E_X[X \mid Y]] = E_Y[q \lfloor \sqrt{Y} \rfloor] \\
= q E[\lfloor \sqrt{Y} \rfloor] \leq q \sqrt{E[Y]} = q \sqrt{np}
\]
Martingales

Definition

A sequence of random variables $Z_0, Z_1, \ldots$ is a martingale with respect to the sequence $X_0, X_1, \ldots$ if for all $n \geq 0$ the following hold:

1. $Z_n$ is a function of $X_0, X_1, \ldots, X_n$;
2. $\mathbb{E}[|Z_n|] < \infty$;
3. $\mathbb{E}[Z_{n+1}|X_0, X_1, \ldots, X_n] = Z_n$;

Definition

A sequence of random variables $Z_0, Z_1, \ldots$ is a martingale when it is a martingale with respect to itself, that is

1. $\mathbb{E}[|Z_n|] < \infty$;
2. $\mathbb{E}[Z_{n+1}|Z_0, Z_1, \ldots, Z_n] = Z_n$;
A series of fair games ($\mathbb{E}[\text{gain}] = 0$), not necessarily independent.

Game 1: bet $1$.

Game $i > 1$: bet $2^i$ if won in round $i - 1$; bet $i$ otherwise.

$X_i =$ amount won in $i$th game. ($X_i < 0$ if $i$th game lost).

$Z_i =$ total winnings at end of $i$th game.
Example

\( X_i = \) amount won in \( i \)th game. \((X_i < 0 \text{ if } i \text{th game lost})\).

\( Z_i = \) total winnings at end of \( i \)th game.

\( Z_1, Z_2, \ldots \) is martingale with respect to \( X_1, X_2, \ldots \)

\( E[X_i] = 0 \).

\( E[Z_i] = \sum_{j=1}^{i} E[X_j] = 0 < \infty \).

\( E[Z_{i+1} | X_1, X_2, \ldots, X_i] = Z_i + E[X_{i+1}] = Z_i \).
I play series of fair games (win with probability 1/2).

Game 1: bet $1.

Game $i > 1$: bet $2^i$ if I won in round $i - 1$; bet $i$ otherwise.

$X_i =$ amount won in $i$th game. ($X_i < 0$ if $i$th game lost).

$Z_i =$ total winnings at end of $i$th game.

Assume that (before starting to play) I decide to quit after $k$ games: what are my expected winnings?
Lemma

Let $Z_0, Z_1, Z_2, \ldots$ be a martingale with respect to $X_0, X_1, \ldots$. For any fixed $n$,

$$E_{X[1:n]}[Z_n] = E[Z_0].$$

$(X[1 : i] = X_1, \ldots, X_i)$

Proof.

Since $Z_i$ is a martingale $E_{X_i}[Z_i|X_0, X_1, \ldots, X_{i-1}] = Z_{i-1}$. Then

$$E_{X[1:i-1]}[Z_{i-1}] = E_{X[1:i-1]}[E_{X_i}[Z_i|X_0, X_1, \ldots, X_{i-1}]] = E_{X[1:i]}[Z_i]$$

Thus,

$$E_{X[1:n]}[Z_n] = E_{X[n-1]}[Z_{n-1}] = \ldots, = E[Z_0]$$
Gambling Strategies

I play series of fair games (win with probability $1/2$).

Game 1: bet $1$.

Game $i > 1$: bet $2^i$ if I won in round $i - 1$; bet $i$ otherwise.

$X_i =$ amount won in $i$th game. ($X_i < 0$ if $i$th game lost).

$Z_i =$ total winnings at end of $i$th game.

Assume that (before starting to gamble) we decide to quit after $k$ games: what are my expected winnings?

$E[Z_k] = E[Z_1] = 0$. 
A Different Strategy

Same gambling game. What happens if I:

- play a random number of games?
- decide to stop only when I have won $1000?
Stopping Time

**Definition**

A non-negative, integer *random variable* $T$ is a *stopping time* for the sequence $Z_0, Z_1, \ldots$ if the event “$T = n$” depends only on the value of random variables $Z_0, Z_1, \ldots, Z_n$.

Intuition: corresponds to a strategy for determining when to stop a sequence based only on values seen so far.

In the gambling game:

- *first time I win 10 games in a row*: is a stopping time;
- *the last time when I win*: is not a stopping time.
Consider again the gambling game: let $T$ be a stopping time.

$Z_i =$ total winnings at end of $i$th game.

What are my winnings at the stopping time, i.e. $E[Z_T]$?

Fair game: $E[Z_T] = E[Z_0] = 0$?

“$T =$first time my total winnings are at least $1000” is a stopping time, and $E[Z_T] > 1000$...
Martingale Stopping Theorem

**Theorem**

If $Z_0, Z_1, \ldots$ is a martingale with respect to $X_1, X_2, \ldots$ and if $T$ is a stopping time for $X_1, X_2, \ldots$ then

$$E[Z_T] = E[Z_0]$$

whenever one of the following holds:

- there is a constant $c$ such that, for all $i$, $|Z_i| \leq c$;
- $T$ is bounded;
- $E[T] < \infty$, and there is a constant $c$ such that $E[|Z_{i+1} - Z_i| \mid X_1, \ldots, X_i] < c$. 


Proof of Martingale Stopping Theorem (Sketch)

We need to show that $E[|Z_T|] < \infty$. So we can use

$E[Z_T] = E[Z_0] + \sum_{i \leq T} E[E[(Z_i - Z_{i-1}) | X_1, \ldots, X_{i-1}]]$

- there is a constant $c$ such that, for all $i$, $|Z_i| \leq c$ - the sum is bounded.
- $T$ is bounded - the sum has finite number of elements.
- $E[T] < \infty$, and there is a constant $c$ such that $E[|Z_{i+1} - Z_i| | X_1, \ldots, X_i] < c$

$$
E[|Z_T|] \leq E[|Z_0|] + \sum_{i=1}^{\infty} E[E[|Z_{i+1} - Z_i| | X_1, \ldots, X_i] 1_{i \leq T}]

\leq E[|Z_0|] + c \sum_{i=1}^{\infty} Pr(T \geq i)

\leq E[|Z_0|] + cE[T] < \infty
$$
We play a sequence of fair game with the following stopping rules:

1. \( T \) is chosen from distribution with finite expectation:
   \( \mathbb{E}[Z_T] = \mathbb{E}[Z_0] \).

2. \( T \) is the first time we made $1000: \mathbb{E}[T] \) is unbounded.

3. We double until the first win. \( \mathbb{E}[T] = 2 \) but
   \( \mathbb{E}[|Z_{i+1} - Z_i| | X_1, \ldots, X_i] \) is unbounded.
Example: The Gambler’s Ruin

- Consider a sequence of independent, fair 2-player gambling games.
- In each round, each player wins or loses $1 with probability $\frac{1}{2}$.
- $X_i =$ amount won by player 1 on $i$th round.
  - If player 1 has lost in round $i$: $X_i < 0$.
- $Z_i =$ total amount won by player 1 after $i$th rounds.
  - $Z_0 = 0$.
- Game ends when one player runs out of money
  - Player 1 must stop when she loses net $\ell_1$ dollars ($Z_t = -\ell_1$)
  - Player 2 terminates when she loses net $\ell_2$ dollars ($Z_t = \ell_2$).
- $q =$ probability game ends with player 1 winning $\ell_2$ dollars.
Example: The Gambler’s Ruin

- \( T \) = first time player 1 wins \( \ell_2 \) dollars or loses \( \ell_1 \) dollars.
  - \( T \) is a stopping time for \( X_1, X_2, \ldots \).
- \( Z_0, Z_1, \ldots \) is a martingale.
  - \( Z_i \)'s are bounded.
- Martingale Stopping Theorem: \( \mathbb{E}[Z_T] = \mathbb{E}[Z_0] = 0 \).

\[
\mathbb{E}[Z_T] = q\ell_2 - (1 - q)\ell_1 = 0
\]

\[
q = \frac{\ell_1}{\ell_1 + \ell_2}
\]
Example: A Ballot Theorem

- Candidate A and candidate B run for an election.
  - Candidate A gets \( a \) votes.
  - Candidate B gets \( b \) votes.
  - \( a > b \).
- Votes are counted in *random order*:
  - chosen from all permutations on all \( n = a + b \) votes.
- What is the probability that A is always ahead in the count?
Example: A Ballot Theorem

- $S_i =$ number of votes A is leading by after $i$ votes counted
  - If A is trailing: $S_i < 0$.
  - $S_n = a - b$.

- For $0 \leq k \leq n - 1$: $X_k = \frac{S_{n-k}}{n-k}$.

- Consider $X_0, X_1, \ldots, X_n$.
  - This sequence goes backward in time!

$$
\mathbb{E}[X_k|X_0, X_1, \ldots, X_{k-1}] = ?
$$
Example: A Ballot Theorem

\[ \mathbb{E}[X_k|X_0, X_1, \ldots, X_{k-1}] = ? \]

- Conditioning on \( X_0, X_1, \ldots, X_{k-1} \): equivalent to conditioning on \( S_n, S_{n-1}, \ldots, S_{n-k+1} \).
- \( a_i = \) number of votes for \( A \) after first \( i \) votes are counted.
- \((n - k + 1)\)th vote: random vote among these first \( n - k + 1 \) votes.

\[
S_{n-k} = \begin{cases} 
S_{n-k+1} + 1 & \text{if } (n - k + 1)\text{th vote is for } B \\
S_{n-k+1} - 1 & \text{if } (n - k + 1)\text{th vote is for } A 
\end{cases}
\]

\[
S_{n-k} = \begin{cases} 
S_{n-k+1} + 1 & \text{with prob. } \frac{n-k+1-a_{n-k+1}}{n-k+1} \\
S_{n-k+1} - 1 & \text{with prob. } \frac{a_{n-k+1}}{n-k+1}
\end{cases}
\]
\[
E[S_{n-k}|S_{n-k+1}] = (S_{n-k+1} + 1) \frac{n - k + 1 - a_{n-k+1}}{(n - k + 1)} \\
+ (S_{n-k+1} - 1) \frac{a_{n-k+1}}{(n - k + 1)} \\
= S_{n-k+1} \frac{n - k}{n - k + 1}
\]

(Since \(2a_{n-k+1} - n - k + 1 = S_{n-k+1}\))

\[
E[X_k|X_0, X_1, \ldots, X_{k-1}] = E \left[ \frac{S_{n-k}}{n-k} \middle| S_n, \ldots, S_{n-k+1} \right] \\
= S_{n-k+1} \frac{n - k}{n - k + 1} \\
= X_{k-1}
\]

\[\implies X_0, X_1, \ldots, X_n \text{ is a martingale.}\]
Example: A Ballot Theorem

\[ T = \begin{cases} 
\min\{k : X_k = 0\} & \text{if such } k \text{ exists} \\
n - 1 & \text{otherwise} 
\end{cases} \]

- \( T \) is a stopping time.
- \( T \) is bounded.
- Martingale Stopping Theorem:

\[
E[X_T] = E[X_0] = E[S_n] = \frac{a - b}{n} = \frac{a - b}{a + b}.
\]

Two cases:

1. \( A \) leads throughout the count.
2. \( A \) does not lead throughout the count.
1. A leads throughout the count.
For $0 \leq k \leq n - 1$: $S_{n-k} > 0$, then $X_k > 0$.

\[ T = n - 1. \]

\[ X_T = X_{n-1} = S_1. \]

A gets the first vote in the count: $S_1 = 1$, then $X_T = 1$.

2. A does not lead throughout the count.
For some $k$: $S_k = 0$. Then $X_k = 0$.

\[ T = k < n - 1. \]

\[ X_T = 0. \]
Example: A Ballot Theorem

Putting all together:

1. A leads throughout the count: $X_T = 1$.
2. A does not lead throughout the count: $X_T = 0$

$$E[X_T] = \frac{a - b}{a + b} = 1 \times \Pr(\text{Case 1}) + 0 \times \Pr(\text{Case 2}).$$

That is

$$\Pr(\text{A leads throughout the count}) = \frac{a - b}{a + b}.$$
A Different Gambling Game

Two stages:

1. roll one die; let $X$ be the outcome;
2. roll $X$ standard dice; your gain $Z$ is the sum of the outcomes of the $X$ dice.

What is your expected gain?
Wald’s Equation

Theorem

Let $X_1, X_2, \ldots$ be nonnegative, independent, identically distributed random variables with distribution $X$. Let $T$ be a stopping time for this sequence. If $T$ and $X$ have bounded expectation, then

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

Note that $T$ is not independent of $X_1, X_2, \ldots$. Corollary of the martingale stopping theorem.
Proof

For \( i \geq 1 \), let \( Z_i = \sum_{j=1}^{i} (X_j - E[X]) \).

The sequence \( Z_1, Z_2, \ldots \) is a martingale with respect to \( X_1, X_2, \ldots \). \( E[Z_1] = 0 \), \( E[T] < \infty \), and since \( X_i \) are nonnegative

\[
E[|Z_{i+1} - Z_i| \mid X_1, \ldots, X_i] = E[|X_{i+1} - E[X]|] \leq 2E[X].
\]

Hence we can apply the martingale stopping theorem to compute

\[
E[Z_T] = E[Z_1] = 0.
\]

We now find

\[
0 = E[Z_T] = E \left[ \sum_{j=1}^{T} (X_j - E[X]) \right] = E \left[ \sum_{j=1}^{T} X_j - T E[X] \right] \\
= E \left[ \sum_{j=1}^{T} X_j \right] - E[T] \cdot E[X] = 0,
\]
A Different Gambling Game

Two stages:
1. roll one die; let $X$ be the outcome;
2. roll $X$ standard dice; your gain $Z$ is the sum of the outcomes of the $X$ dice.

What is your expected gain?

$Y_i =$ outcome of $i$th die in second stage.

$$E[Z] = E \left[ \sum_{i=1}^{X} Y_i \right].$$

$X$ is a stopping time for $Y_1, Y_2, \ldots$.

By Wald’s equation:

$$E[Z] = E[X]E[Y_i] = \left( \frac{7}{2} \right)^2.$$
Example: a \( k \)-run

- We flip a fair coin until we get a consecutive sequence of \( k \) HEADs.
- What’s the expected number of times we flip the coin.
- A SWITCH is a HEAD followed by a TAIL.
- Let \( X_1 \) be the number of flips till \( k \) HEADs or the first SWITCH.
- Let \( X_i \) be the number of flips following the \( i - 1 \) SWITCH till \( k \) HEADs or the next SWITCH (\( X_i \) includes the last HEAD or TAIL).
- Let \( T \) be the first \( i \) with \( k \) HEADs.

\[
\begin{align*}
\mathbb{E}[X_i] &= \sum_{j \geq 1} j2^{-j} + \sum_{j=1}^{k-1} j2^{-j} + (k - 1)2^{-(k-1)} \\
\mathbb{E}[T] &= 2^{k-1}
\end{align*}
\]

- The expected number of coin flips is \( \mathbb{E}[X_i]\mathbb{E}[T] \).
Let $X_i$ be the number of flips following the $i - 1$ SWITCH till $k$ HEADs or the next SWITCH ($X_i$ includes the last HEAD or TAIL).

Let $T$ be the first $i$ with $k$ HEADs.

$X_i =$ number of flips till (including) first HEAD + up to $k - 2$ HEADs followed by a TAIL, or $k - 1$ HEADS.

$$\mathbb{E}[X_i] = \sum_{j \geq 1} j2^{-j} + \sum_{j=1}^{k-1} j2^{-j} + (k - 1)2^{-(k-1)}$$

The probability that $X_i$ ends with $k$ HEADS is $2^{-(k-1)}$ - sequence of $k - 1$ HEADS following the first one.

$$\mathbb{E}[T] = 2^{k-1}$$

The expected number of coin flips is $\mathbb{E}[X_i] \mathbb{E}[T]$.
Hoeffding’s Bound

Theorem

Let $X_1, \ldots, X_n$ be \textbf{independent} random variables with $E[X_i] = \mu_i$ and $\Pr(B_i \leq X_i \leq B_i + c_i) = 1$, then

$$\Pr \left( \left| \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i \right| \geq \epsilon \right) \leq 2e^{-\frac{2\epsilon^2}{\sum_{i=1}^{n} c_i^2}}$$

Do we need independence?
**Tail Inequalities**

**Theorem (Azuma-Hoeffding Inequality)**

Let $Z_0, Z_1, \ldots, Z_n$ be a martingale (with respect to $X_1, X_2, \ldots$) such that $|Z_k - Z_{k-1}| \leq c_k$. Then, for all $t \geq 0$ and any $\lambda > 0$,

$$\Pr(|Z_t - Z_0| \geq \lambda) \leq 2e^{-\lambda^2/(2 \sum_{k=1}^{t} c_k^2)} .$$

The following corollary is often easier to apply.

**Corollary**

Let $X_0, X_1, \ldots$ be a martingale such that for all $k \geq 1$,

$$|X_k - X_{k-1}| \leq c .$$

Then for all $t \geq 1$ and $\lambda > 0$,

$$\Pr(|X_t - X_0| \geq \lambda c \sqrt{t}) \leq 2e^{-\lambda^2/2} .$$
Theorem (Azuma-Hoeffding Inequality)

Let $Z_0, Z_1, \ldots$, be a martingale with respect to $X_0, X_1, X_2, \ldots$, such that

$$B_k \leq Z_k - Z_{k-1} \leq B_k + c_k,$$

for some constants $c_k$ and for some random variables $B_k$ that may be functions of $X_0, X_1, \ldots, X_{k-1}$. Then, for any $t \geq 0$ and $\lambda > 0$,

$$\Pr(|Z_t - Z_0| \geq \lambda) \leq 2e^{-2\lambda^2/(\sum_{k=1}^{t} c_k^2)}.$$
Proof
Let \( X^k = X_0, \ldots, X_k \) and \( Y_i = Z_i - Z_{i-1} \).

Since \( E[Z_i \mid X^{i-1}] = Z_{i-1} \),
\[
E[Y_i \mid X^{i-1}] = E[Z_i - Z_{i-1} \mid X^{i-1}] = 0 .
\]

Since \( \Pr(B_i \leq Y_i \leq B_i + c_i \mid X^{i-1}) = 1 \), by Hoeffding’s Lemma:
\[
E[e^{\beta Y_i} \mid X^{i-1}] \leq e^{\beta^2 c_i^2 / 8} .
\]

Lemma

(Hoeffding’s Lemma) Let \( X \) be a random variable such that \( \Pr(X \in [a, b]) = 1 \) and \( E[X] = 0 \). Then for every \( \lambda > 0 \),
\[
E[e^{\lambda X}] \leq e^{\lambda^2 (a - b)^2 / 8} .
\]
Proof of the Lemma

Since \( f(x) = e^{\lambda x} \) is a convex function, for any \( \alpha \in (0, 1) \) and \( x \in [a, b] \),

\[
f(X) \leq \alpha f(a) + (1 - \alpha) f(b) .
\]

Thus, for \( \alpha = \frac{b - x}{b - a} \in (0, 1) \),

\[
e^{\lambda x} \leq \frac{b - x}{b - a} e^{\lambda a} + \frac{x - a}{b - a} e^{\lambda b} .
\]

Taking expectation, and using \( \mathbf{E}[X] = 0 \), we have

\[
\mathbf{E} \left[ e^{\lambda X} \right] \leq \frac{b}{b - a} e^{\lambda a} + \frac{a}{b - a} e^{\lambda b} \leq e^{\lambda^2 (b-a)^2 / 8} .
\]
Proof of Azuma-Hoeffding Inequality

\[ \mathbb{E} \left[ e^{\beta Y_i} \mid X^{i-1} \right] \leq e^{\beta^2 c_i^2 / 8} . \]

\[
\begin{align*}
\mathbb{E}_{X^n} \left[ e^\beta \sum_{i=1}^n Y_i \right] &= \mathbb{E}_{X^{n-1}} \left[ \mathbb{E}_{X^n} \left[ e^\beta \sum_{i=1}^n Y_i \mid X^{n-1} \right] \right] \\
&= \mathbb{E}_{X^n} \left[ e^\beta \sum_{i=1}^{n-1} Y_i \mathbb{E}_{X_{n-1}} \left[ e^\beta Y_n \mid X^{n-1} \right] \right] \\
&\leq e^{\beta^2 c_n^2 / 8} \mathbb{E}_{X^{n-1}} \left[ e^\beta \sum_{i=1}^{n-1} Y_i \right] \\
&\leq e^{\beta^2 \sum_{i=1}^n c_i^2 / 8}
\end{align*}
\]
\[ E[e^\beta \sum_{i=1}^n Y_i] \leq e^{\beta^2} \sum_{i=1}^n c_i^2 / 8 \]

\[ \Pr(Z_t - Z_0 \geq \lambda) = \Pr\left( \sum_{i=1}^t Y_i \geq \lambda \right) \leq \frac{E[e^\beta \sum_{i=1}^t Y_i]}{e^{\beta \lambda}} \]

\[ \leq e^{-\lambda \beta} e^{\beta^2} \sum_{i=1}^t c_i^2 / 8 \]

\[ \leq 2e^{-2\lambda^2} / (\sum_{k=1}^t c_k^2) \]

For \( \beta = \frac{4\lambda}{\sum_{i=1}^t c_i^2} \).

\[ \Pr(|Z_t - Z_0| \geq \lambda) \leq 2e^{-2\lambda^2} / (\sum_{k=1}^t c_k^2) \]
Example

Assume that you play a sequence of $n$ fair games, where the bet $b_i$ in game $i$ depends on the outcome of previous games. Let $B = \max_i b_i$. The probability of winning or losing more than $\lambda$ is bounded by

$$\Pr(|Z_n| \geq \lambda) \leq 2e^{-2\lambda^2/nB^2}$$

$$\Pr(|Z_n| \geq \lambda B \sqrt{n}) \leq 2e^{-2\lambda^2}$$

$$\Pr \left( |Z_n| \geq \lambda \sqrt{\sum_{i=1}^{n} b_i^2} \right) \leq 2e^{-2\lambda^2}$$
Doob Martingale

Let $X_1, X_2, \ldots, X_n$ be sequence of random variables. Let $Y = f(X_1, \ldots, X_n)$ be a random variable with $E[|Y|] < \infty$.

For $i = 0, 1, \ldots, n$, let

$$
Z_0 = E[Y] = E_{X[1,n]} f(X_1, \ldots, X_n)
$$

$$
Z_i = E_{X[i+1,n]} [Y|X_1 = x_1, X_2 = x_2, \ldots, X_i = x_i]
$$

$$
Z_n = E[Y|X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n] = f(x_1, \ldots, x_n)
$$

Theorem

$Z_0, Z_1, \ldots, Z_n$ is martingale with respect to $X_1, X_2, \ldots, X_n$. 
Proof

We use:

<table>
<thead>
<tr>
<th>Fact</th>
</tr>
</thead>
</table>

\[ Y = f(X_1, \ldots, X_n), \quad Z_0 = \mathbb{E}[Y], \]
\[ Z_i = \mathbb{E}_{X[i+1,n]}[Y|X_1 = x_1, \ldots, X_i = x_i], \]
\[ Z_1, Z_2, \ldots, Z_n \] is a martingale if
\[
\mathbb{E}_{X[i+1]}[Z_{i+1}|X_1 = x_1, \ldots, X_i = x_i] = Z_i
\]

\[
\mathbb{E}_{X[i+1]}[Z_{i+1}|x_1, x_2, \ldots, x_i] = \mathbb{E}_{X[i+1]}[\mathbb{E}_{X[i+2,n]}[Y|X_1, \ldots, X_{i+1}]|x_1, \ldots, x_i] \\
= \mathbb{E}_{X[i+1,n]}[Y|x_1, x_2, \ldots, x_i] \\
= Z_i .
\]
Simple Example

\[ Y = f(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i, \]

\( X_i \)'s are independent and distributed uniform \( U[0, 1] \).

\[
Z_0 = E[Y] = E_{X[1,n]}[f(X_1, \ldots, X_n)] = E[\sum_{i=1}^{n} X_i] = n/2
\]

\[
Z_i = E_{X[i+1,n]}[Y|x_1, \ldots, x_i]
\]

\[
= \sum_{j=1}^{i} x_j + E[\sum_{j=i}^{n} X_i] = \sum_{j=1}^{i} x_j + (n - i)/2
\]

\[
Z_n = E[Y|x_1, \ldots, x_n] = f(x_1, \ldots, x_n) = \sum_{j=1}^{n} x_j
\]

\[
E_{X_{i+1}}[Z_{i+1}|x_1, \ldots, x_i] = E_{X_{i+1}} \left[ \sum_{j=1}^{i+1} X_j + \frac{n - i - 1}{2} \right] |x_1, \ldots, x_i
\]

\[
= \sum_{i-1}^{i} x_i + \frac{n - i}{2} = Z_i
\]
Example: Polya’s Urn

- Start with $M$ balls, $R$ red, $M - R$ blue.
- Repeat $n$ times: We pick a ball uniformly at random. If Red we add a red ball, else we add a blue ball.
- $X_i = 1$ if we add a red ball in step $i$, else $X_i = 0$
- We want to estimate $S_n(R/M) = \sum_{i=1}^{n} X_i = f(X_1, \ldots, X_n)$
- **Claim:** $E[S_n(R/M)] = nR/M$. Proof by induction on $t$, that $E[S_t] = tR/M$.

$$E[S_{t+1} | S_t] = S_t + \frac{R + S_t}{M + t}$$

$$E[S_{t+1}] = E[E[S_{t+1} | S_t]] = E\left[S_t + \frac{R + S_t}{M + t}\right]$$

$$= t\frac{R}{M} + \frac{R + tR/M}{M + t} = (t + 1)\frac{R}{M}$$
Example: Polyga’s Urn

Start with $M$ balls, $R$ red, $M - R$ blue. Repeat $n$ times: We pick a ball uniformly at random. If Red we add a red ball, else we add a blue ball. $X_i = 1$ if added a red ball in step $i$, else $X_i = 0$, $S_n(R/M) = \sum_{i=1}^{n} X_i$, and $E[S_n(R/M)] = nR/M$

Let $Z_i = E[S_n | X_1 = x_1, \ldots, X_i = x_i]$. We prove that $Z_1, \ldots, Z_n$ is a martingale.

$$Z_i = E[S_n | X_1 = x_1, \ldots, X_i = x_i] = \sum_{j=1}^{i} x_j + E[S_{n-i} \left( \frac{R + \sum_{j=1}^{i} x_j}{M + i} \right)]$$

$$= \sum_{j=1}^{i} x_j + (n - i) \frac{R + \sum_{j=1}^{i} x_j}{M + i}$$

$$E[Z_{i+1} | X_1, \ldots, X_i] = E[E[S_n | X_1, X_2, \ldots, X_{i+1}] | X_1 = x_1, \ldots, X_i = x_i]$$

$$= E \left[ \sum_{j=1}^{i} x_j + X_{i+1} + S_{n-i-1} \left( \frac{R + \sum_{j=1}^{i} x_j + X_{i+1}}{M + i + 1} \right) \right]$$
\[ Z_i = \mathbb{E}[S_n \mid X_1 = x_1, \ldots, X_i = x_i] = \sum_{j=1}^i x_j + (n - i) \frac{R + \sum_{j=1}^i x_j}{M + i} \]

\[ \mathbb{E}[Z_{i+1} \mid X_1, \ldots, X_i] = \mathbb{E} \left[ \sum_{j=1}^i x_j + X_{i+1} + S_{n-i-1} \left( \frac{R + \sum_{j=1}^i x_j + X_{i+1}}{M + i + 1} \right) \right] \]

\[ = \mathbb{E} \left[ \sum_{j=1}^i x_j + X_{i+1} + (n - i - 1) \frac{R + \sum_{j=1}^i x_j + X_{i+1}}{M + i + 1} \right] \]

\[ = \sum_{j=1}^i x_j + \frac{R + \sum_{j=1}^i x_j}{M + i} + (n - i - 1) \frac{R + \sum_{j=1}^i x_j + \frac{R + \sum_{j=1}^i x_j}{M + i}}{M + i + 1} \]

\[ = \sum_{j=1}^i x_j + \frac{R + \sum_{j=1}^i x_j}{M + i} + (n - i - 1) \frac{M + i + 1}{M + i} \left( R + \sum_{j=1}^i x_j \right) = Z_i \]
Example: Edge Exposure Martingale

Let $G$ random graph from $G_{n,p}$. Consider $m = \binom{n}{2}$ possible edges in arbitrary order.

$$X_i = \begin{cases} 
1 & \text{if } i\text{th edge is present} \\
0 & \text{otherwise}
\end{cases}$$

$F(G) =$ size of maximum clique in $G$.

$Z_0 = \mathbb{E}[F(G)]$

$Z_i = \mathbb{E}[F(G)|X_1, X_2, \ldots, X_i]$, for $i = 1, \ldots, m$.

$Z_0, Z_1, \ldots, Z_m$ is a Doob martingale.

($F(G)$ could be any finite-valued function on graphs.)
Tail Inequalities: Doob Martingales

Let $X_1, \ldots, X_n$ be sequence of random variables.

Random variable $Y$:
- $Y$ is a function of $X_1, X_2, \ldots, X_n$;
- $E[|Y|] < \infty$.

Let $Z_i = E[Y|X_1, \ldots, X_i], i = 0, 1, \ldots, n$.

$Z_0, Z_1, \ldots, Z_n$ is martingale with respect to $X_1, \ldots, X_n$.

If we can use Azuma-Hoeffding inequality:

$$\Pr(|Z_n - Z_0| \geq \lambda) \leq \ldots.$$

then we have,

$$\Pr(|Y - E[Y]| \geq \lambda) \leq \ldots.$$

We need a bound on $|Z_i - Z_{i-1}|$. 
Theorem

Assume that $f(X_1, X_2, \ldots, X_n)$ satisfies,

$$|f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, y_i, \ldots, x_n)| \leq c_i.$$  

and $X_1, \ldots, X_n$ are independent, then

$$\Pr(|f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)]| \geq \lambda) \leq 2e^{-2\lambda^2/(\sum_{k=1}^{n} c_k^2)}.$$  

[Changing the value of $X_i$ changes the value of the function by at most $c_i$.]
Proof

Define a Doob martingale $Z_0, Z_1, \ldots, Z_n$:

- $Z_0 = \mathbb{E}[f(X_1, \ldots, X_n)] = \mathbb{E}[f(\bar{X})]$  
- $Z_i = \mathbb{E}[f(X_0, \ldots, X_n) \mid X_1, \ldots, X_i] = \mathbb{E}[f(X_i, \ldots, X_n) \mid X^i]$  
- $Z_n = f(X_1, \ldots, X_n) = f(\bar{X})$

We want to prove that this martingale satisfies the conditions of

**Theorem (Azuma-Hoeffding Inequality)**

Let $Z_0, Z_1, \ldots,$ be a martingale with respect to $X_0, X_1, X_2, \ldots$, such that

$$B_k \leq Z_k - Z_{k-1} \leq B_k + c_k,$$

for some constants $c_k$ and for some random variables $B_k$ that may be functions of $X_0, X_1, \ldots, X_{k-1}$. Then, for all $t \geq 0$ and any $\lambda > 0$,

$$\Pr(|Z_t - Z_0| \geq \lambda) \leq 2e^{-2\lambda^2/(\sum_{k=1}^t c_k^2)}.$$
Lemma

If $X_1, \ldots, X_n$ are independent then for some random variable $B_k$,

$$B_k \leq Z_k - Z_{k-1} \leq B_k + c_k,$$

Thus, we need to show

$$Z_k - Z_{k-1} = \mathbb{E}[f(\bar{X}) \mid X^k] - \mathbb{E}[f(\bar{X}) \mid X^{k-1}] .$$

Hence $Z_k - Z_{k-1}$ is bounded above by

$$\sup_x \mathbb{E}[f(\bar{X}) \mid X^{k-1}, X_k = x] - \mathbb{E}[f(\bar{X}) \mid X^{k-1}]$$

and bounded below by

$$\inf_y \mathbb{E}[f(\bar{X}) \mid X^{k-1}, X_k = y] - \mathbb{E}[f(\bar{X}) \mid X^{k-1}] .$$

Thus, we need to show

$$\sup_x \mathbb{E}[f(\bar{X}) \mid X^{k-1}, X_k = x] - \inf_y \mathbb{E}[f(\bar{X}) \mid X^{k-1}, X_k = y] \leq c ,$$
\[ Z_k - Z_{k-1} = \sup_{x,y} E[f(\bar{X}, x) - f(\bar{X}, y) | X^{k-1}] \]

Because the \( X_i \) are independent, the values for \( X_{k+1}, \ldots, X_n \) do not depend on the values of \( X_1, \ldots, X_k \). Hence, for any values \( \bar{z}, x, y \) we have \( \Pr(\bar{z}, x) = \Pr(\bar{z}, y) \), and therefore

\[
\sup_{x,y} E[f(\bar{X}, x) - f(\bar{X}, y) | X_1 = z_1, \ldots, X_{k-1} = z_{k-1}] = \\
\sup_{x,y} \sum_{z_{k+1}, \ldots, z_n} \Pr((X_{k+1} = z_{k+1}) \cap \ldots \cap (X_n = z_n)) \times (f(\bar{z}, x) - f(\bar{z}, y)) .
\]

But

\[ f(\bar{z}, x) - f(\bar{z}, y) \leq c_i \]

and therefore

\[ E[f(\bar{X}, x) - f(\bar{X}, y) | X^{k-1}] \leq c_i \]
Example: Pattern Matching

Given a string and a pattern: is the pattern interesting?

Does it appear more often than is expected in a random string?

Is the number of occurrences of the pattern concentrated around the expectation?
A = \( (a_1, a_2, \ldots, a_n) \) string of characters, each chosen independently and uniformly at random from \( \Sigma \), with \( s = |\Sigma| \).

pattern: \( B = (b_1, \ldots, b_k) \) fixed string, \( b_i \in \Sigma \).

\( F = \) number of occurrences of \( B \) in random string \( A \).

\[ \mathbb{E}[F] = ? \]
A = \((a_1, a_2, \ldots, a_n)\) string of characters, each chosen independently and uniformly at random from \(\Sigma\), with \(m = |\Sigma|\).

pattern: \(B = (b_1, \ldots, b_k)\) fixed string, \(b_i \in \Sigma\).

\(F\) = number occurrences of \(B\) in random string \(S\).

\[E[F] = (n - k + 1) \left(\frac{1}{m}\right)^k\]

Can we bound the deviation of \(F\) from its expectation?
$F = \text{number occurrences of } B \text{ in random string } A.$

$Z_0 = E[F]$

$Z_i = E[F|a_1, \ldots, a_i], \text{ for } i = 1, \ldots, n.$

$Z_0, Z_1, \ldots, Z_n \text{ is a Doob martingale.}$

$Z_n = F.$
\( F \) = number occurrences of \( B \) in random string \( A \).

\( Z_0 = E[F] \)

\( Z_i = E[F|a_1, \ldots, a_i] \), for \( i = 1, \ldots, n \).

\( Z_0, Z_1, \ldots, Z_n \) is a Doob martingale.

\( Z_n = F \).

Each character in \( A \) can participate in no more than \( k \) occurrences of \( B \):

\[
|Z_i - Z_{i+1}| \leq k .
\]

Azuma-Hoeffding inequality (version 1):

\[
Pr(|F - E[F]| \geq \lambda) \leq 2e^{-\lambda^2/(2nk^2)} .
\]
Application: Balls and Bins

We are throwing $m$ balls independently and uniformly at random into $n$ bins.
Let $X_i =$ the bin that the $i$th ball falls into.
Let $F$ be the number of empty bins after the $m$ balls are thrown.
Then the sequence

$$Z_i = \mathbb{E}[F \mid X_1, \ldots, X_i]$$

is a Doob martingale.
$F = f(X_1, X_2, \ldots, X_m)$ satisfies the Lipschitz condition with bound 1, thus $|Z_{i+1} - Z_i| \leq 1$
We therefore obtain

$$\Pr(|F - \mathbb{E}[F]| \geq \epsilon) \leq 2e^{-2\epsilon^2/m}$$

Here

$$\mathbb{E}[F] = n \left(1 - \frac{1}{n}\right)^m,$$

but we could obtain the concentration result without knowing $\mathbb{E}[F]$. 
Given a random graph $G$ in $G_{n,p}$, the *chromatic number* $\chi(G)$ is the minimum number of colors required to color all vertices of the graph so that no adjacent vertices have the same color.

We use the vertex exposure martingale defined below:

Let $G_i$ be the random subgraph of $G$ induced by the set of vertices $1, \ldots, i$, let $Z_0 = E[\chi(G)]$, and let

$$Z_i = E[\chi(G) \mid G_1, \ldots, G_i].$$

Since a vertex uses no more than one new color, again we have that the gap between $Z_i$ and $Z_{i-1}$ is at most 1.

We conclude

$$\Pr(|\chi(G) - E[\chi(G)]| \geq \lambda \sqrt{n}) \leq 2e^{-2\lambda^2}.$$

This result holds even without knowing $E[\chi(G)]$. 
Example: Edge Exposure Martingale

Let $G$ random graph from $G_{n,p}$. Consider $m = \binom{n}{2}$ possible edges in arbitrary order.

$$X_i = \begin{cases} 1 & \text{if } i\text{th edge is present} \\ 0 & \text{otherwise} \end{cases}$$

$F(G) =$ size maximum clique in $G$.

$$Z_0 = \mathbb{E}[F(G)]$$

$$Z_i = \mathbb{E}[F(G) | X_1, X_2, \ldots, X_i], \text{ for } i = 1, \ldots, m.$$ 

$Z_0, Z_1, \ldots, Z_m$ is a Doob martingale.

($F(G)$ could be any finite-valued function on graphs.)