Hoeffding's Bound

Theorem

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

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

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. $\mathbf{E}[|Z_n|] < \infty$;
3. $\mathbf{E}[Z_{n+1}|X_0, X_1, \ldots, X_n] = Z_n$;

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

Conditional Expectation

Definition

$$\mathbf{E}[Y \mid Z = z] = \sum_y y \Pr(Y = y \mid Z = z) \ ,$$

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

Lemma

For any random variables X and Y,

$$\mathbf{E}[X] = \mathbf{E}_Y[\mathbf{E}_X[X \mid Y]] = \sum_y \Pr(Y = y)\mathbf{E}[X \mid Y = y] \ ,$$

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

Lemma

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

$$\mathbf{E}[X] = \mathbf{E}_Y[\mathbf{E}_X[X \mid Y]] = \sum_y \Pr(Y = y)\mathbf{E}[X \mid Y = y] ,$$

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

Proof.

$$\begin{aligned}
& \sum_y \Pr(Y = y)\mathbf{E}[X \mid Y = y] \\
= & \sum_y \Pr(Y = y) \sum_x x \Pr(X = x \mid Y = y) \\
= & \sum_x \sum_y x \Pr(X = x \mid Y = y) \Pr(Y = y) \\
= & \sum_x \sum_y x \Pr(X = x \cap Y = y) = \sum_x x \Pr(X = x) = \mathbf{E}[X].
\end{aligned}$$

□

Example

Consider a two phase game:

- Phase I: roll one die. Let X be the outcome.
- Phase II: Flip X fair coins, let Y be the number of HEADs.
- You receive a dollar for each HEAD.

Y is distributed $B(X, \frac{1}{2})$,

$$\mathbf{E}[Y \mid X = a] = \frac{a}{2}$$

$$\mathbf{E}[Y] = \sum_{i=1}^{6} \mathbf{E}[Y \mid X = i] \Pr(X = i)$$

$$= \sum_{i=1}^{6} \frac{i}{2} \Pr(X = i) = \frac{7}{4}$$

Conditional Expectation as a Random variable

Definition

The expression $\mathbf{E}[Y \mid Z]$ is a random variable $f(Z)$ that takes on the value $\mathbf{E}[Y \mid Z = z]$ when $Z = z$.

Consider the outcome of rolling two dice $X_1, X_2, X = X_1 + X_2$.

$$\mathbf{E}[X \mid X_1] = \sum_x x \Pr(X = x \mid X_1) = \sum_{x=X_1+1}^{X_1+6} x \cdot \frac{1}{6} = X_1 + \frac{7}{2} \ .$$

Consider the two phase game

$$\mathbf{E}[Y \mid X] = \frac{X}{2}$$

If $\mathbf{E}[Y \mid Z]$ is a random variable, it has an expectation.

Theorem

$$\mathbf{E}[Y] = \mathbf{E}[\mathbf{E}[Y \mid Z]] \ .$$

$$\mathbf{E}[X \mid X_1] = X_1 + \frac{7}{2} \ .$$

Thus

$$\mathbf{E}[\mathbf{E}[X \mid X_1]] = \mathbf{E}\left[X_1 + \frac{7}{2}\right] = \frac{7}{2} + \frac{7}{2} = 7 \ .$$

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. $\mathbf{E}[|Z_n|] < \infty$;
3. $\mathbf{E}[Z_{n+1}|X_0, X_1, \ldots, X_n] = Z_n$;

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

Martingale Example

A series of fair games ($\mathbf{E}[\text{gain}] = 0$), 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 ith game. ($X_i < 0$ if ith game lost).

$Z_i =$ total winnings at end of ith game.

Example

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

$Z_i =$ total winnings at end of ith game.

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

$\mathbf{E}[X_i] = 0.$

$\mathbf{E}[Z_i] = \sum \mathbf{E}[X_i] = 0 < \infty.$

$\mathbf{E}[Z_{i+1} | X_1, X_2, \ldots, X_i] = Z_i + \mathbf{E}[X_{i+1}] = Z_i.$

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 ith game. ($X_i < 0$ if ith game lost).

$Z_i =$ total winnings at end of ith 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,

$$\mathbf{E}_{X[1:n]}[Z_n] = \mathbf{E}[Z_0] \quad .$$

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

Proof.

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

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

But

$$\mathbf{E}_{X[1:i-1]}[\mathbf{E}_{X_i}[Z_i | X_0, X_1, \ldots, X_{i-1}]] = \mathbf{E}_{X[1:i]}[Z_i] \quad .$$

Thus,

$$\mathbf{E}_{X[1:n]}[Z_n] = \mathbf{E}_{X[n-1]}[Z_{n-1}] = \ldots, = \mathbf{E}[Z_0] \quad .$$

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 ith game. ($X_i < 0$ if ith game lost).

$Z_i =$ total winnings at end of ith game.

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

$\mathbf{E}[Z_k] = \mathbf{E}[Z_1] = 0$.

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 ith game.

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

Fair game: $\mathbf{E}[Z_T] = \mathbf{E}[Z_0] = 0$?

"$T =$first time my total winnings are at least $\$1000$" is a stopping time, and $\mathbf{E}[Z_T] > 1000$...

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*

$$\mathbf{E}[Z_T] = \mathbf{E}[Z_0]$$

whenever one of the following holds:

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

Proof of Martingale Stopping Theorem (Sketch)

We need to show that $E[|Z_T|] < \infty$. So we can use $E[Z_T] = E[Z_0] + \sum_{i \le T} E[E[(Z_i - Z_{i-1}) \mid X_1, \ldots, X_{i-1}]]$

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

$$
\begin{aligned}
E[|Z_T|] &\le E[|Z_0|] + \sum_{i=1}^{\infty} \mathbf{E}[\mathbf{E}\left[|Z_{i+1} - Z_i| \middle| X_1, \ldots, X_i\right] \mathbf{1}_{i \le T}] \\
&\le E[|Z_0|] + c \sum_{i=1}^{\infty} Pr(T \ge i) \\
&\le E[|Z_0|] + cE[T] < \infty
\end{aligned}
$$

Martingale Stopping Theorem Applications

We play a sequence of fair game with the following stopping rules:

1. T is chosen from distribution with finite expectation: $\mathbf{E}[Z_T] = \mathbf{E}[Z_0]$.
2. T is the first time we made $1000: $\mathbf{E}[T]$ is unbounded.
3. We double until the first win. $\mathbf{E}[T] = 2$ but $\mathbf{E}\left[|Z_{i+1} - Z_i| \middle| X_1, \ldots, X_i\right]$ 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 ith round.
  - If player 1 has lost in round i: $X_i < 0$.
- $Z_i =$ total amount won by player 1 after ith rounds.
  - $Z_0 = 0$.
- Game ends when one player runs out of money
  - Player 1 must stop when she loses net ℓ_1 dollars ($Z_t = -\ell_1$)
  - Player 2 terminates when she loses net ℓ_2 dollars ($Z_t = \ell_2$).
- $q =$ probability game ends with player 1 winning ℓ_2 dollars.

Example: The Gambler's Ruin

- $T =$ first time player 1 wins ℓ_2 dollars or loses ℓ_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: $\mathbf{E}[Z_T] = \mathbf{E}[Z_0] = 0$.

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

$$q = \frac{\ell_1}{\ell_1 + \ell_2}$$

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!

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

Example: A Ballot Theorem

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

- 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.

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

$$S_{n-k} = \begin{cases} S_{n-k+1} + 1 & \text{with prob. } \frac{n-k+1-a_{n-k+1}}{n-k+1} \\ S_{n-k+1} - 1 & \text{with prob. } \frac{a_{n-k+1}}{n-k+1} \end{cases}$$

$$
\begin{aligned}
\mathbf{E}[S_{n-k}|S_{n-k+1}] &= (S_{n-k+1}+1)\frac{n-k+1-a_{n-k+1}}{(n-k+1)} \\
&+ (S_{n-k+1}-1)\frac{a_{n-k+1}}{(n-k+1)} \\
&= S_{n-k+1}\frac{n-k}{n-k+1}
\end{aligned}
$$

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

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

$\implies \quad X_0, X_1, \ldots, X_n$ 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:

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

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: $X_T = 0$

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

That is

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

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

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

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

Proof

For $i \geq 1$, let $Z_i = \sum_{j=1}^{i}(X_j - \mathbf{E}[X])$.
The sequence $Z_1, Z_2, \ldots$ is a martingale with respect to $X_1, X_2, \ldots$.
$\mathbf{E}[Z_1] = 0$, $\mathbf{E}[T] < \infty$, and since X_i are nonnegative

$$\mathbf{E}\left[|Z_{i+1} - Z_i| \;\middle|\; X_1, \ldots, X_i\right] = \mathbf{E}[|X_{i+1} - \mathbf{E}[X]|] \leq 2\mathbf{E}[X] \ .$$

Hence we can apply the martingale stopping theorem to compute

$$\mathbf{E}[Z_T] = \mathbf{E}[Z_1] = 0 \ .$$

We now find

$$\begin{aligned} 0 &= \mathbf{E}[Z_T] = \mathbf{E}\left[\sum_{j=1}^{T}(X_j - \mathbf{E}[X])\right] = \mathbf{E}\left[\sum_{j=1}^{T} X_j - T\mathbf{E}[X]\right] \\ &= \mathbf{E}\left[\sum_{j=1}^{T} X_j\right] - \mathbf{E}[T] \cdot \mathbf{E}[X] = 0, \end{aligned}$$

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 ith die in second stage.

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

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

By Wald's equation:

$$\mathbf{E}[Z] = \mathbf{E}[X]\mathbf{E}[Y_i] = \left(\frac{7}{2}\right)^2 .$$

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

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

- The expected number of coin flips is $\mathbf{E}[X_i]E[T]$

- 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

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

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

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

- The expected number of coin flips is $\mathbf{E}[X_i]E[T]$

Hoeffding's Bound

Theorem

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

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

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 $\lambda > 0$,

$$\Pr\left(|X_t - X_0| \geq \lambda c\sqrt{t}\right) \leq 2e^{-\lambda^2/2} \ .$$

Tail Inequalities: A More General Form

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 $X^k = X_0, \dots, X_k$ and $Y_i = Z_i - Z_{i-1}$.

Since $\mathbf{E}[Z_i \mid X^{i-1}] = Z_{i-1}$,

$$\mathbf{E}[Y_i \mid X^{i-1}] = \mathbf{E}[Z_i - Z_{i-1} \mid X^{i-1}] = 0 \ .$$

Since $\Pr(B_i \le Y_i \le B_i + c_i \mid X^{i-1}) = 1$, by Hoeffding's Lemma:

$$\mathbf{E}[e^{\beta Y_i} \mid X^{i-1}] \le e^{\beta^2 c_i^2/8} \ .$$

Lemma

(Hoeffding's Lemma) Let X be a random variable such that $\Pr(X \in [a,b]) = 1$ and $\mathbf{E}[X] = 0$. Then for every $\lambda > 0$,

$$\mathbf{E}[e^{\lambda X}] \le e^{\lambda^2 (a-b)^2/8}.$$

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

$$\mathbf{E}\left[e^{\beta Y_i} \;\middle|\; X^{i-1}\right] \leq e^{\beta^2 c_i^2/8} \quad .$$

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

$$\mathbf{E}[e^{\beta \sum_{i=1}^{n} Y_i}] \leq e^{\beta^2 \sum_{i=1}^{n} c_i^2/8}$$

$$\begin{aligned}
\Pr(Z_t - Z_0 \geq \lambda) &= \Pr\left(\sum_{i=1}^{t} Y_i \geq \lambda\right) \leq \frac{\mathbf{E}[e^{\beta \sum_{i=1}^{t} Y_i}]}{e^{\beta\lambda}} \\
&\leq e^{-\lambda\beta} e^{\beta^2 \sum_{i=1}^{t} c_i^2/8} \\
&\leq 2e^{-2\lambda^2/(\sum_{k=1}^{t} c_k^2)},
\end{aligned}$$

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 λ 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 $\mathbf{E}[|Y|] < \infty$.

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

$$\begin{aligned} Z_0 &= \mathbf{E}[Y] = \mathbf{E}_{X[1,n]} f(X_1, \ldots, X_n] \\ Z_i &= \mathbf{E}_{X[i+1,n]}[Y | X_1 = x_1, X_2 = x_2, \ldots, X_i = x_i] \\ Z_n &= \mathbf{E}[Y | X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n] = f(x_1, \ldots, x_n) \end{aligned}$$

Theorem

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

Proof

We use:

Fact

$\mathbf{E}[\mathbf{E}[V|U,W]|W] = \mathbf{E}[V|W]$.

$Y = f(X_1,\ldots,X_n)$, $Z_0 = \mathbf{E}[Y]$,
$Z_i = \mathbf{E}_{X[i+1,n]}[Y|X_1 = x_1,\ldots,X_i = x_i]$,

$Z_1, Z_2.\ldots, Z_n$ is a martingale if

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

$$\begin{aligned}
\mathbf{E}_{X_{i+1}}[Z_{i+1}|x_1,x_2,\ldots,x_i] &= \mathbf{E}_{X_{i+1}}[\mathbf{E}_{X[i+2,n]}[Y|X_1,\ldots,X_{i+1}]|x_1,...,x_i] \\
&= \mathbf{E}_{X[i+1,n]}[Y|x_1,x_2,\ldots,x_i] \\
&= Z_i \quad .
\end{aligned}$$

Simple Example

$Y = f(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$,
X_i's are independent and distributed uniform $U[0, 1]$.

$$\begin{aligned}
Z_0 &= \mathbf{E}[Y] = \mathbf{E}_{X[1,n]} f(X_1, \ldots, X_n)] = \mathbf{E}[\sum_{i=1}^{n} X_i] = n/2 \\
Z_i &= \mathbf{E}_{X[i+1,n]}[Y | x_1, \ldots, x_i] \\
&= \sum_{j=1}^{i} x_j + \mathbf{E}[\sum_{j=i}^{n} X_i] = \sum_{j=1}^{i} x_j + (n-i)/2 \\
Z_n &= \mathbf{E}[Y | x_1, \ldots, x_n] = f(x_1, \ldots, x_n) = \sum_{j=1}^{n} x_j
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}_{X_{i+1}}[Z_{i+1} | x_1, \ldots, x_i] &= \mathbf{E}_{X_{i+1}}\left[\sum_{j=1}^{i+1} X_i + \frac{n-i-1}{2} | x_1, \ldots, x_i\right] \\
&= \sum_{j=1}^{i} x_i + \frac{n-i}{2} = Z_i
\end{aligned}$$

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:** $\mathbf{E}[S_n(R/M)] = nR/M$.
  Proof by induction on t, that $\mathbf{E}[S_t] = tR/M$.

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

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

$$= t\frac{R}{M} + \frac{R + tR/M}{M + t} = (t + 1)\frac{R}{M}$$

Example: 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 added a red ball in step i, else $X_i = 0$, $S_n(R/M) = \sum_{i=1}^{n} X_i$, and $\mathbf{E}[S_n(R/M)] = nR/M$

Let $Z_i = \mathbf{E}[S_n \mid X_1 = x_1, \dots, X_i = x_i]$. We prove that $Z_1, \dots, Z_n$ is a martingale.

$$
\begin{aligned}
Z_i &= \mathbf{E}[S_n \mid X_1 = x_1, \dots, X_i = x_i] = \sum_{j=1}^{i} x_j + E[S_{n-i}(\frac{R + \sum_{j=1}^{i} x_j}{M+i})] \\
&= \sum_{j=1}^{i} x_j + (n-i)\frac{R + \sum_{j=1}^{i} x_j}{M+i}
\end{aligned}
$$

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

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

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

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

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

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

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

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.

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

$Z_i = \mathbf{E}[F(G) | X_1, X_2, \ldots, X_i]$, 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 $X_1, X_2, \ldots, X_n$;
- $\mathbf{E}[|Y|] < \infty$.

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

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

If we can use Azuma-Hoeffding inequality:

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

then we have,

$$\Pr(|Y - \mathbf{E}[Y]| \geq \lambda) \leq ......$$

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

McDiarmid Bound

Theorem

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

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

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

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

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

Proof

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

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

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,

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

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

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

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

and bounded below by

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

Thus, we need to show

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

$$Z_k - Z_{k-1} \;=\; \sup_{x,y} \mathbf{E}[f(\bar{X}, x) - f(\bar{X}, y) \mid X^{k-1}].$$

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

$$\sup_{x,y} \mathbf{E}[f(\bar{X}, x) - f(\bar{X}, y) \mid X_1 = z_1, \ldots, X_{k-1} = z_{k-1}] =$$

$$\sup_{x,y} \sum_{z_{k+1},\ldots,z_n} \Pr\big((X_{k+1} = z_{k+1}) \cap \ldots \cap (X_n = z_n)\big) * \big(f(\bar{z}, x) - f(\bar{z}, y)\big) \; .$$

But

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

and therefore

$$\mathbf{E}[f(\bar{X}, x) - f(\bar{X}, y) \mid X^{k-1}] \leq c_i$$

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 Σ, 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.

$$\mathbf{E}[F] = ?$$

$A = (a_1, a_2, \ldots, a_n)$ string of characters, each chosen independently and uniformly at random from Σ, 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.

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

Can we bound the deviation of F from its expectation?

$F=$ number occurrences of B in random string A.

$Z_0 = \mathbf{E}[F]$

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

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

$Z_n = F$.

$F=$ number occurrences of B in random string A.

$Z_0 = \mathbf{E}[F]$

$Z_i = \mathbf{E}[F | a_1, \ldots, a_i]$, 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:

$$|Z_i - Z_{i+1}| \leq k \quad .$$

Azuma-Hoeffding inequality (version 1):

$$\Pr(|F - \mathbf{E}[F]| \geq \lambda) \leq 2e^{-\lambda^2/(2nk^2)} \quad .$$

Application: Balls and Bins

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

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

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

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

Here

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

but we could obtain the concentration result without knowing $\mathbf{E}[F]$.

Application: Chromatic Number

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 $Z_0 = \mathbf{E}[\chi(G)]$, and let

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

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

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

This result holds even without knowing $\mathbf{E}[\chi(G)]$.

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.

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

$Z_i = \mathbf{E}[F(G) | X_1, X_2, \ldots, X_i]$, 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.)