Introduction

Some preliminaries

Examples

A concentration inequality

An application to chromatic numbers
Definition (Martingales)

Let $X = (X_n)_{n \geq 0}$ and $Y = (Y_n)_{n \geq 0}$ be two sequences of random variables.
Suppose that for some Borel function $g$, $X_n = g(Y_n, Y_{n-1}, \ldots, Y_0)$. $X$ is said to
be a martingale with respect to $Y$ if

$$
E(|X_n|) < \infty \quad \text{for all } n,
$$

$$
E(X_n|Y_0, \ldots, Y_{n-1}) = X_{n-1} \quad a.s \quad \text{for all } n, \ n \geq 1.
$$

Martingales are stochastic processes that 'tend to remain where they are' as time
passes.
Some examples

**Random Walk** Let $X_n$, $n \geq 1$ be independent r.v. with zero mean. Let $X_0 = 0$. Let $S_n = \sum_{k\leq n} X_k$ for $n \geq 0$. Then $S = (S_n)_n$ is a martingale with respect to $X = (X_n)_n$.

**Learning about a r.v.** Let $Z$ be a random variable with finite mean. Let $X = (X_n)$ be a sequence of random variables. Let $M_n = E(Z|X_0, \ldots, X_n)$. Then $M = (M_n)_n$ is a martingale with respect to $X$. 
Let $G_{n,p}$ be an Erdos-Renyi Random Graph: we have $n$ nodes, each pair of nodes is connected independently with probability $p$.

**Edge exposure martingale** Let $Z_i$ be the indicator variable of whether edge $i$ is present in the graph. Let $A = f(Z_1, \ldots, Z_{n \choose 2})$ be a graph property and $X_i = E(A|Z_1, \ldots Z_i)$.

**Vertex exposure martingale** Let $Z_1 = 0$ and for $i > 1$ let $Z_i \in \{0, 1\}^{i-1}$ be a vector of indicators of whether edges between vertex $i$ and vertices $j < i$ are present in the graph. Let $A = f(Z_1, \ldots, Z_n)$ be a graph property and $X_i = E(A|Z_1, \ldots Z_i)$. 
Some preliminaries
All non-deterministics statements hold a.s. Again, let $Z, W, Y_0, \ldots, Y_n$ be r.vs.

- If $Z = g(Y_0, \ldots, Y_n)$, $E(ZW|Y_0, \ldots, Y_n) = ZE(W|Y_0, \ldots, Y_n)$.
- If $Z$ is independent of $Y_0, \ldots, Y_n$, $E(Z|Y_0, \ldots, Y_n) = E(Z)$.
- $E(E(Z|Y_0, \ldots, Y_n)|Y_0, \ldots, Y_{n-1}) = E(Z|Y_0, \ldots, Y_{n-1})$.
- If $\phi$ is convex, $E(\phi(Z)|Y_0 \ldots Y_n) \geq \phi(E(Z|Y_0 \ldots Y_n))$.
- $E(a_1Z_1 + a_2Z_2|Y_0, \ldots, Y_n) = a_1E(Z_1|Y_0, \ldots, Y_n) + a_2E(Z_2|Y_0, \ldots, Y_n)$.
- $E(E(Z|Y_0 \ldots Y_n)) = E(Z)$.
- If $Z_1 \leq Z_2$ then $E(Z_1|Y_0, \ldots, Y_n) \leq E(Z_2|Y_0, \ldots, Y_n)$.
Examples
Random Walk

Example (Random Walk)

Let’s check \((S_n)_n\) is a martingale.

- \(E(|S_n|) \leq \sum_{k \leq n} E(|X_k|) < \infty\).
- \(E(S_n|X_0, \ldots, X_{n-1}) = E(S_{n-1}|X_0, \ldots, X_{n-1}) = S_{n-1}\), due to the \(X_n\)s being independent.
Learning about a r.v.

Martingales of the form $M_n = E(Z|X_0,\ldots,X_n)$ are also called Doob’s martingales.

Example

Let’s check $(M_n)_n$ is a martingale.

- $E(|M_n|) = E(|E(Z|X_0,\ldots,X_n)|) \leq E(E(|Z||X_0,\ldots,X_n)) = E(|Z|) < \infty$.
- $E(M_n|X_0,\ldots,X_{n-1}) = E(E(Z|X_0,\ldots,X_n)|X_0,\ldots,X_{n-1}) = E(Z|X_0,\ldots,X_{n-1}) = M_{n-1}$. 

Let $X = (X_n)_n$ be a martingale.

- How and when can we bound $P(|X_n - X_0| > \varepsilon)$?
- Let $A \subset \mathbb{R}$. Let $T = \inf\{n : X_n \in A\}$. What’s $E(T)$? What can we say about $X_T$?
- When does there exist a random variable $X_\infty$ such that $X_n \to X_\infty$? In what sense?
A concentration inequality
Theorem (Azuma-Hoeffding)

Let \((X_n)_n\) be a martingale with respect to \((Y_n)_n\). Let \(D_n = X_n - X_{n-1}\). Assume there exists constants \((d_n)_n\) such that \(|D_n| \leq d_n\) almost surely. Then for any \(t > 0\)

\[
P (X_n - X_0 \geq t) \leq \exp \left( \frac{-t^2}{2 \sum_{k \leq n} d_k^2} \right).
\]
Proof of Azuma-Hoeffding

For simplicity assume $X_0 = 0$. Using Markov’s inequality we get that for all $\theta, t > 0$,

$$P (X_n \geq t) \leq \frac{E (\exp (\theta X_n))}{\exp (\theta t)}.$$

Now

$$E (\exp (\theta X_n)) = E (E (\exp (\theta X_n)) | Y_0, \ldots, Y_{n-1}).$$

Note that

$$\exp (\theta X_n) = \exp (\theta (X_{n-1} + D_n)) = \exp (\theta X_{n-1}) \exp (\theta D_n).$$
Proof of Azuma-Hoeffding

Since $X_{n-1} = g(Y_{n-1}, \ldots, Y_0)$,

$$E(\exp(\theta X_n)) = E \left( E(\exp(\theta X_n)|Y_0, \ldots, Y_{n-1}) \right)$$
$$= E \left( E(\exp(\theta X_{n-1}) \exp(\theta D_n)|Y_0, \ldots, Y_{n-1}) \right)$$
$$= E \left( \exp(\theta X_{n-1}) E(\exp(\theta D_n)|Y_0, \ldots, Y_{n-1}) \right).$$

Suppose we can show that $E(\exp(\theta D_k)|Y_0, \ldots, Y_{k-1}) \leq \exp(\theta^2 d_k^2 / 2)$. Then

$$E(\exp(\theta X_n)) \leq E(\exp(\theta X_{n-1})) \exp(\theta^2 d_{n-1}^2 / 2) \leq \cdots \leq \prod_{k \leq n} \exp(\theta^2 d_k^2 / 2)$$

and

$$P(X_n \geq t) \leq \prod_{k \leq n} \exp(\theta^2 d_k^2 / 2) \exp(-t\theta).$$
Minimizing the RHS of

\[ P(X_n \geq t) \leq \prod_{k \leq n} \exp(\theta^2 d_k^2 / 2) \exp(-t\theta) \]

over \( \theta \) gives that the optimal \( \theta \) is \( t / \sum_{k \leq n} d_k^2 \) and we get the bound

\[ P(X_n \geq t) \leq \exp \left( \frac{-t^2}{2 \sum_{k \leq n} d_k^2} \right). \]

To show \( E(\exp(\theta D_k))|Y_0, \ldots, Y_{k-1}) \leq \exp(\theta^2 d_k^2 / 2) \) we will need a lemma.
Lemma (Hoeffding’s Lemma)

Let $V$ be a r.v. and $Z$ a random vector such that $V \in [a, b]$ a.s., where $a < 0 < b$. Assume $E(V|Z) = 0$. Then $E(\exp(\theta V)|Z) \leq \exp((\theta(b-a))^2/8)$.

Proof.

$\exp(\theta x)$ is convex. Take $x \in [a, b]$ and let $\lambda = (b-x)/(b-a) \in [0, 1]$. Note that $\lambda a + (1-\lambda)b = x$. Then

$$\exp(\theta x) = \exp(\lambda \theta a + (1-\lambda)\theta b) \leq \lambda \exp(\theta a) + (1-\lambda) \exp(\theta b)$$

$$= \frac{b-x}{b-a} \exp(\theta a) + \frac{x-a}{b-a} \exp(\theta b)$$

Hence

$$E(\exp(\theta V)|Z) \leq \exp(\theta a)E\left(\frac{b-V}{b-a}|Z\right) + \exp(\theta b)E\left(\frac{V-a}{b-a}|Z\right)$$

$$= \exp(\theta a) \frac{b}{b-a} + \exp(\theta b) \frac{-a}{b-a}$$
Let $p = -a/(b - a)$. Then

$$E(\exp(\theta V) | Z) \leq \exp(-\theta p(b - a))(1 - p) + \exp(\theta(1 - p)(b - a))p$$

$$= \exp(-\theta p(b - a))((1 - p) + p \exp(\theta(b - a)))$$

$$= \exp(-pu + \log(1 - p + p \exp(u))) \leq \exp(u^2/8),$$

where $u = \theta(b - a)$. 
By symmetry, we also get

\[ P(X_n - X_0 \leq -t) \leq \exp \left( \frac{-t^2}{2 \sum_{k \leq n} d_k^2} \right). \]

and

\[ P(|X_n - X_0| \geq t) \leq 2 \exp \left( \frac{-t^2}{2 \sum_{k \leq n} d_k^2} \right). \]
McDiarmid’s Inequality

Theorem (McDiarmid’s Inequality)

Let \( X_1, \ldots, X_n \) be independent random elements of \( \mathcal{X}_1, \ldots, \mathcal{X}_n \) respectively. Let \( \mathcal{X} = \prod_{i=1}^{n} \mathcal{X}_i \). Suppose that \( f : \mathcal{X} \to \mathbb{R} \) satisfies that: if \( x \) and \( x' \) differ only on the \( k \)-th coordinate then \( |f(x) - f(x')| \leq \sigma_k \). Let \( Y = f(X_1, \ldots, X_n) \). Then

\[
P \left( |Y - E(Y)| \geq t \right) \leq 2 \exp \left( \frac{-t^2}{2 \sum_{k \leq n} \sigma_k^2} \right).
\]
Proof.

Let $Y_k = E(Y|X_1, \ldots, X_k)$ and $Y_0 = E(Y)$. We will show that $Y_k$ satisfies the hypothesis of Azuma-Hoeffding. Write $X_{i:j} = (X_i, \ldots, X_j)$, $x_{i:j} = (x_i, \ldots, x_j)$.

\[
Y_k = E(Y|X_1, \ldots, X_k)
= \sum_{x_{k+1:n}} f(X_{1:k}, x_{k+1:n}) P(X_{k+1:n} = x_{k+1:n}|X_{1:k})
= \sum_{x_{k+1:n}} f(X_{1:k}, x_{k+1:n}) P(X_{k+1:n} = x_{k+1:n}).
\]
Proof.

\[ Y_{k-1} = \sum_{x_{k:n}} f(X_{1:k-1}, x_{k:n}) P(X_{k:n} = x_{k:n}) \]

\[ = \sum_{x_{k+1:n}} \sum_{x'_k} f(X_{1:k-1}, x'_k, x_{k+1:n}) P(X_k = x'_k) P(X_{k+1:n} = x_{k+1:n}) \]
Proof.

Thus

\[ |Y_k - Y_{k-1}| \leq \sum_{x_{k+1:n}} f(X_{1:k}, x_{k+1:n}) - \sum_{x_k'} f(X_{1:k-1}, x_k', x_{k+1:n}) P(X_k = x_k') |P(X_{k+1:n} = x_{k+1:n}) \]

\[ \leq \sum_{x_{k+1:n}} \sigma_k P(X_{k+1:n} = x_{k+1:n}) \leq \sigma_k. \]
An application to chromatic numbers
Consider an Erdos-Renyi random graph $G_{n,p}$. Let $\chi$ be its chromatic number. Let $Z_1 = 0$ and for $i > 1$ let $Z_i \in \{0, 1\}^{i-1}$ be a vector of indicators of whether edges between vertex $i$ and vertices $j < i$ are present in the graph. Note that

$$\chi = f(Z_1, \ldots, Z_n),$$

for some function $f$. 
If we modify $Z_i$ by adding edges incident to $i$, we can maintain a proper coloring by eventually adding a new color for $i$, hence the chromatic number increases by at most 1; if we modify $Z_i$ by removing edges incident to $i$, the chromatic number cannot decrease by more than 1.

Hence we have (Shamir and Spencer 87’)

$$P(|\chi - E(\chi)| \geq t) \leq 2 \exp\left(-\frac{t^2}{2n}\right).$$

In particular, deviations of $X$ from $E(X)$ by an order greater than $\sqrt{n}$ are unlikely.