Randomized Algorithms

南京大学

尹一通
Definition:
A sequence of random variables $X_0, X_1, \ldots$ is a **martingale** if for all $i > 0$,
\[
E[X_i | X_0, \ldots, X_{i-1}] = X_{i-1}
\]

\[
\forall x_0, x_1, \ldots, x_{i-1},
E[X_i | X_0 = x_0, X_1 = x_1, \ldots, X_{i-1} = x_{i-1}] = x_{i-1}
\]
Azuma’s Inequality:

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

$$|X_k - X_{k-1}| \leq c_k,$$

Then

$$\Pr[|X_n - X_0| \geq t] \leq 2 \exp\left(-\frac{t^2}{2 \sum_{k=1}^{n} c_k^2}\right).$$

- For a sequence, if in each step:
  - averagely no change to the current value (martingale),
  - no big jump,
  - the final does not deviate far from the initial.
Azuma’s Inequality:
Let $X_0, X_1, \ldots$ be a martingale such that, for all $k \geq 1$,
$$|X_k - X_{k-1}| \leq c_k,$$
Then
$$\Pr[|X_n - X_0| \geq t] \leq 2 \exp\left(-\frac{t^2}{2 \sum_{k=1}^{n} c_k^2}\right).$$

1. Represent total difference as sum of step-wise differences.
   Let $Y_i = X_i - X_{i-1}$. \quad $X_n - X_0 = \sum_{i=1}^{n} Y_i$

2. Apply Markov’s inequality to the moment generating function.
   $$\Pr[\sum_{i=1}^{n} Y_i \geq t] = \Pr[e^{\lambda \sum_{i=1}^{n} Y_i} \geq e^{\lambda t}] \leq \frac{\mathbb{E}[e^{\lambda \sum_{i=1}^{n} Y_i}]}{e^{\lambda t}}$$

3. Bound the moment generating function.
   by martingale property & convexity of MGF
Generalization

**Definition:**

$Y_0, Y_1, \ldots$ is a martingale with respect to $X_0, X_1, \ldots$ if, for all $i \geq 0$,

- $Y_i$ is a function of $X_0, X_1, \ldots, X_i$;
- $\mathbb{E}[Y_{i+1} | X_0, \ldots, X_i] = Y_i$. 
Azuma’s Inequality (general version):

Let $Y_0, Y_1, \ldots$ be a martingale with respect to $X_0, X_1, \ldots$ such that, for all $k \geq 1$,

$$|Y_k - Y_{k-1}| \leq c_k,$$

Then

$$\Pr[|Y_n - Y_0| \geq t] \leq 2 \exp\left(-\frac{t^2}{2 \sum_{k=1}^{n} c_k^2}\right).$$
Definition (Doob sequence):
The Doob sequence of a function $f$ with respect to a sequence $X_1, \ldots, X_n$ is

$$ Y_i = \mathbb{E}[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i] $$

$$ Y_0 = \mathbb{E}[f(X_1, \ldots, X_n)] \quad \text{-------} \quad Y_n = f(X_1, \ldots, X_n) $$
Doob sequence:
\[ Y_i = \mathbb{E}[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i] \]

Doob sequence is a martingale:
\[ \mathbb{E}[Y_i \mid X_1, \ldots, X_{i-1}] = Y_{i-1} \]

Proof:
\[
\begin{align*}
\mathbb{E}[Y_i \mid X_1, \ldots, X_{i-1}] &= \mathbb{E}[\mathbb{E}[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i] \mid X_1, \ldots, X_{i-1}] \\
&= \mathbb{E}[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_{i-1}] \\
&= Y_{i-1}
\end{align*}
\]
Doob Sequence

randomized by

\[ f(1, \epsilon, \epsilon, \epsilon, \epsilon, \epsilon, \epsilon) \]

averaged over
Doob Sequence

randomized by $f(1, 0, \ldots, \delta, \ldots)$

averaged over
Doob Sequence

\[ f(1, 0, 0, \ldots) \]

randomized by

averaged over
Doob Sequence

\[ f(1, 0, 0, 1, \cdot, \cdot) \]

randomized by

averaged over
Doob Sequence

$f((1, 0, 0, 1, 0, 0))$ randomized by averaged over
Doob Sequence

$f(1, 0, 0, 1, 0, 1)$ randomized by
Doob Martingale
Graph parameter: $f(G)$

example: chromatic #, components, diameter ...

numbering all vertex-pairs: $1, 2, 3, \ldots, \binom{n}{2}$

$I_j = \begin{cases} 
1 & \text{edge } j \in G \\
0 & \text{edge } j \notin G 
\end{cases}$

$Y_i = \mathbb{E}[f(G) | I_1, \ldots, I_i]$  

$Y_0 = \mathbb{E}[f(G)] \quad \longrightarrow \quad Y_{\binom{n}{2}} = f(G)$
$G(n, p)$

Graph parameter: $f(G)$

**example**: chromatic #, components, diameter ... 

numbering all vertices: $1, 2, 3, \ldots, n$

$X_i$: subgraph of $G$ induced by the first $i$ vertices

$Y_i = \mathbb{E}[f(G) | X_1, \ldots, X_i]$

$Y_0 = \mathbb{E}[f(G)] \quad \longrightarrow \quad Y_n = f(G)$
Martingales induced by a random graph

- **Edge exposure martingale:**
  
  \[ I_j \text{ indicates the } j\text{th edge} \]

  \[ Y_i = \mathbb{E}[f(G) \mid I_1, \ldots, I_i] \]

- **Vertex exposure martingale:**
  
  \[ X_i = G([i]) \]

  \[ Y_i = \mathbb{E}[f(G) \mid X_1, \ldots, X_i] \]
martingale \( X_0, X_1, X_2, \ldots \)
\[
E[X_i \mid X_0, X_1, \ldots, X_{i-1}] = X_{i-1}
\]

generalization

martingale \( Y_0, Y_1, Y_2, \ldots \)
\[
w.r.t. \ X_0, X_1, X_2, \ldots
\]
\[
Y_i = f(X_0, X_1, \ldots, X_i)
\]
\[
E[Y_i \mid X_0, X_1, \ldots, X_{i-1}] = Y_{i-1}
\]

edge-exposure martingale
vertex-exposure martingale

special cases in random graphs

Doob martingale
\[
Y_i = E[f(X_0, X_1, \ldots, X_n) \mid X_0, X_1, \ldots, X_{i-1}]
\]
Hoeffding’s Inequality:

Let $X = \sum_{i=1}^{n} X_i$, where $X_1, \ldots, X_n$ are independent random variables with

$$a_i \leq X_i \leq b_i.$$ 

Then

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq 2 \exp \left(- \frac{t^2}{2 \sum_{i=1}^{n} (b_i - a_i)^2} \right)$$
Hoeffding’s Inequality:

Let \( X = \sum_{i=1}^{n} X_i \), where \( X_1, \ldots, X_n \) are independent random variables with

\[
    a_i \leq X_i \leq b_i.
\]

Then

\[
    \Pr[|X - \mathbb{E}[X]| \geq t] \leq 2 \exp\left( -\frac{t^2}{2 \sum_{i=1}^{n} (b_i - a_i)^2} \right)
\]

Proof:  \( X \) is a function (sum) of \( X_1, \ldots, X_n \).

Doob martingale:  \( Y_i = \mathbb{E}[X | X_1, \ldots, X_i] \)

\[
    |Y_i - Y_{i-1}| = |X_i - \mathbb{E}[X_i]| \leq b_i - a_i
\]

Azuma  \rightarrow  Done!
The Power of Doob + Azuma

- For a function of (dependent) random variables: $f(X_1, \ldots, X_n)$

- **Doob martingale:**
  
  $$Y_i = \mathbb{E}[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i]$$

  $$Y_0 = \mathbb{E}[f(X_1, \ldots, X_n)] \quad Y_n = f(X_1, \ldots, X_n)$$

- If the **differences** $|Y_i - Y_{i-1}|$ are bounded,

- **Azuma**

  $$|f(X_1, Y_0), iX_n)| is tightly concentrated to its mean
The method of bounded differences:

Let $X = (X_1, \ldots, X_n)$ and let $f$ be a function of $X_0, X_1, \ldots, X_n$ satisfying that, for all $1 \leq i \leq n$,

$$|E[f(X) \mid X_1, \ldots, X_i] - E[f(X) \mid X_1, \ldots, X_{i-1}]| \leq c_i,$$

Then

$$\Pr[|f(X) - E[f(X)]| \geq t] \leq 2 \exp \left(-\frac{t^2}{2 \sum_{i=1}^{n} c_i^2}\right).$$
The method of bounded differences:

Let $X = (X_1, \ldots, X_n)$ and let $f$ be a function of $X_0, X_1, \ldots, X_n$ satisfying that, for all $1 \leq i \leq n$,

$$\left| E[f(X) \mid X_1, \ldots, X_i] - E[f(X) \mid X_1, \ldots, X_{i-1}] \right| \leq c_i,$$

Then (Azuma)

$$\Pr \left[ |f(X) - E[f(X)]| \geq t \right] \leq 2 \exp \left( -\frac{t^2}{2 \sum_{i=1}^{n} c_i^2} \right).$$

Doob martingale: $Y_i = E[f(X) \mid X_1, \ldots, X_i]$
The method of bounded differences:

Let $X = (X_1, \ldots, X_n)$ and let $f$ be a function of $X_0, X_1, \ldots, X_n$ satisfying that, for all $1 \leq i \leq n$,

$$|E[f(X) \mid X_1, \ldots, X_i] - E[f(X) \mid X_1, \ldots, X_{i-1}]| \leq c_i,$$

Then hard to check!

$$\Pr[\|f(X) - E[f(X)]\| \geq t] \leq 2 \exp \left( -\frac{t^2}{2 \sum_{i=1}^n c_i^2} \right).$$
Lipschitz Condition:

\( f(x_1, \ldots, x_n) \) satisfies the **Lipschitz condition** with constants \( c_i, 1 \leq i \leq n \), if

\[
\left| f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n) \right| \leq c_i.
\]
Average-case:

\[ |\mathbb{E}[f(X) | X_1, \ldots, X_i] - \mathbb{E}[f(X) | X_1, \ldots, X_{i-1}]| \leq c_i, \]

Worst-case:

**Lipschitz Condition:**

\( f(x_1, \ldots, x_n) \) satisfies the **Lipschitz condition** with constants \( c_i, 1 \leq i \leq n \), if

\[
|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)| \leq c_i.
\]
The method of bounded differences:

Let $X = (X_1, \ldots, X_n)$ be $n$ independent random variables and let $f$ be a function satisfying the Lipschitz condition with constants $c_i$, $1 \leq i \leq n$. Then

$$\Pr \left[ \left| f(X) - \mathbb{E}[f(X)] \right| \geq t \right] \leq 2 \exp \left( - \frac{t^2}{2 \sum_{i=1}^{n} c_i^2} \right).$$
The method of bounded differences:

Let $X = (X_1, \ldots, X_n)$ be $n$ independent random variables and let $f$ be a function satisfying the Lipschitz condition with constants $c_i, 1 \leq i \leq n$. Then

$$\Pr[|f(X) - \mathbb{E}[f(X)]| \geq t] \leq 2 \exp\left(-\frac{t^2}{2 \sum_{i=1}^{n} c_i^2}\right).$$

Proof:

Lipschitz condition + independence → bounded averaged differences
Applications of the Method

• A function $f$ of independent random variable:
  $$X_1, X_2, \ldots, X_n$$

• Lipschitz condition of $f$:
  • changes of any variable makes little change to the value of $f$.  

• $\Rightarrow f(X_1, X_2, \ldots, X_n)$ is tightly concentrated to its mean.
Occupancy Problem

- $m$-balls-into-$n$-bins:
- number of empty bins?

$X_i = \begin{cases} 
1 & \text{bin } i \text{ is empty} \\
0 & \text{otherwise} 
\end{cases}$

$X = \sum_{i=1}^{n} X_i$

$E[X] = \sum_{i=1}^{n} E[X_i] = n \left(1 - \frac{1}{n}\right)^m$

deviation: $\Pr[|X - E[X]| \geq t] \leq ?$

$X_i$ are dependent
Occupancy Problem

- $m$-balls-into-$n$-bins:
- number of empty bins?

$Y_j$: the bin of ball $j$ (Independent!)

$$X = f(Y_1, \ldots, Y_m) = |[n] - \{Y_1, \ldots, Y_m\}|$$

Lipschitz:
changing any $Y_j$ can change $X$ for at most 1

Pr[| $X - E[X]$ | ≥ $t \sqrt{m}$] ≤ $2e^{-t^2/2}$
Pattern Matching

- a random string of length \( n \),
- a pattern of length \( k \),
- \# of matched substrings?

alphabet \( \Sigma \)

\[ |\Sigma| = m \]

a fixed pattern:

\[ \pi \in \Sigma^k \]

uniform & independent:

\[ X_1, \ldots, X_n \in \Sigma \]

\[ Y : \text{#substrings } \pi \text{ in } (X_1, \ldots, X_n) \]

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

Deviation?
Pattern Matching

- a random string of length $n$, 
- a pattern of length $k$,
- # of matched substrings?

uniform & independent: $X_1, \ldots, X_n \in \Sigma$

$$Y = f(X_1, \ldots, X_n)$$

changing any $X_i$ changes $f$ for at most $k$

$$\Pr[|Y - \mathbb{E}[Y]| \geq tk\sqrt{n}] \leq 2e^{-t^2/2}$$
Metric Embedding

\[(X, d_X) \rightarrow (Y, d_Y)\]

low-distortion: For a small \(\alpha \geq 1\)

\[\forall x_1, x_2 \in X, \quad \frac{1}{\alpha} d_X(x_1, x_2) \leq d_Y(\phi(x_1), \phi(x_2)) \leq \alpha d_X(x_1, x_2)\]
Why Embedding?

- Some problems are easier to solve in simple metrics:
  - TSP in trees.
  - “Curse of dimensionality”
    - proximity search;
    - learning;
    - due to volume explosion.
Dimension Reduction

In Euclidian space, it is always possible to embed a set of $n$ points in arbitrary dimension to $O(\log n)$ dimension with constant distortion.

**Johnson-Lindenstrauss Theorem:**

For any $0 < \epsilon < 1$, for any set $V$ of $n$ points in $\mathbb{R}^d$, there is a map $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^k$ with $k = O(\ln n)$, such that $\forall u, v \in V$,

$$(1 - \epsilon)\|u - v\|^2 \leq \|\phi(u) - \phi(v)\|^2 \leq (1 + \epsilon)\|u - v\|^2$$
Johnson-Lindenstrauss Theorem:

For any $0 < \epsilon < 1$, for any set $V$ of $n$ points in $\mathbb{R}^d$, there is a map $\phi : \mathbb{R}^d \to \mathbb{R}^k$ with $k = O(\ln n)$, such that $\forall u, v \in V$,

$$(1 - \epsilon)\|u - v\|^2 \leq \|\phi(u) - \phi(v)\|^2 \leq (1 + \epsilon)\|u - v\|^2$$

- $\phi(v) = Av$.
- $A$ is a random projection matrix.
Random Projection

Random $k \times d$ matrix $A$:

- Projection onto a uniform random subspace. 
  (Johnson-Lindenstrauss) 
  (Dasgupta-Gupta)

- i.i.d. Gaussian entries. 
  (Indyk-Motiwani)

- i.i.d. -1/+1 entries. 
  (Achlioptas)

rows: $A_1, A_2, \ldots, A_k$.

random orthogonal unit vectors \( \in \mathbb{R}^d \)