Lecture 7: "Martingales"

Sotiris Nikoletseas
Associate Professor

Computer Engineering and Informatics Department
2015 - 2016

Summary of previous lecture

1. The Janson Inequality
2. Example - Triangle-free sparse Random Graphs
3. Example - Paths of length 3 in $G_{n,p}$

Summary of this lecture

1) Probability theory preliminaries
2) Martingales
3) Example
4) Doob martingales
5) Edge exposure martingale
6) Edge exposure martingale - Example
7) Vertex exposure martingale
8) Azuma's inequality
9) Lipschitz condition
10) Example - Chromatic number
11) Example - Balls and bins

Probability theory preliminaries

If X and Y are discrete random variables then:

1. Joint probability mass function:

$$f(x,y) = \Pr\{X = x \cap Y = y\}$$

2. Conditional Probability:

$$\Pr\{X = x|Y = y\} = \frac{f(x,y)}{\Pr\{Y = y\}} = \frac{f(x,y)}{\sum_x f(x,y)}$$

3. Conditional Expectation:

$$E\big[X|Y = y\big] = \sum_x x \cdot \Pr\{X = x|Y = y\} = \sum_x x \cdot \frac{f(x,y)}{\sum_x f(x,y)}$$

Remark: $E\big[X|Y = y\big] = f(Y)$ is actually a random variable. (depends on the value of Y)

Lemma 1

$$E\big[E[X|Y]\big] = E[X]$$

Proof:
Denote $E[X|Y]$ as a random variable:

$$f(Y) = E[X|Y] = \sum_x x \cdot \frac{f(x,y)}{\Pr\{Y = y\}}$$

Proof of Lemma 1

$$
\begin{aligned}
\Rightarrow E\big[E[X|Y]\big] &= E[f(Y)] = \sum_y f(y)\,\Pr\{Y=y\} \\
&= \sum_y \left(\sum_x x \cdot \frac{f(x,y)}{\Pr\{Y=y\}} \right) \Pr\{Y=y\} \\
&= \sum_y \left(\sum_x x \cdot f(x,y) \right) \\
&= \sum_x x \cdot \left(\sum_y f(x,y) \right) \\
&= \sum_x x \cdot \Pr\{X=x\} \\
&= E[X]
\end{aligned}
$$

□

Martingales

Definition 1

A martingale is a sequence $X_0, X_1, \ldots$ *of random variables so that*

$$\forall i: \quad E[X_i | X_0, \ldots, X_{i-1}] = X_{i-1}$$

Example

- Consider a bin that initially contains b black balls and w white balls.
- We iteratively choose at random a ball from the bin and replace it with c balls of the same color.
- Define random variable X_i which refers to the percentage of black balls after i^{th} iteration.
- The sequence $X_0, X_1, \ldots$ is a martingale.

Proof:
Let as denote that after the $i-1$ iteration there are b_{i-1} black and w_{i-1} white balls in the bin. Thus,

$$X_{i-1} = \frac{b_{i-1}}{b_{i-1} + w_{i-1}}$$

Proof of Example

After the i^{th} iteration:

- **case 1:** The probability of choosing a black ball is

$$X_{i-1} = \frac{b_{i-1}}{b_{i-1} + w_{i-1}}$$

If we choose it and replace it with c black balls the bin will contain:

  - $b_{i-1} + c - 1$ black balls and
  - w_{i-1} white balls

Thus,

$$X_i = \frac{b_{i-1} + c - 1}{b_{i-1} + w_{i-1} + c - 1}$$

Proof of Example

- **case 2:** The probability of choosing a white ball is

$$1 - X_{i-1} = \frac{w_{i-1}}{b_{i-1} + w_{i-1}}$$

  If we choose it and replace it with c white balls the bin will contain :

  - b_{i-1} black balls and
  - $w_{i-1} + c - 1$ white balls

  Thus,

$$X_i = \frac{b_{i-1}}{b_{i-1} + w_{i-1} + c - 1}$$

Proof of Example

$$E[X_i|X_0,\ldots,X_{i-1}] =$$

$$= \frac{b_{i-1}}{b_{i-1}+w_{i-1}} \cdot \frac{b_{i-1}+c-1}{b_{i-1}+w_{i-1}+c-1} + \frac{w_{i-1}}{b_{i-1}+w_{i-1}} \cdot \frac{b_{i-1}}{b_{i-1}+w_{i-1}+c-1}$$

$$= \frac{b_{i-1}\cdot(b_{i-1}+c-1)+w_{i-1}b_{i-1}}{(b_{i-1}+w_{i-1})\cdot(b_{i-1}+w_{i-1}+c-1)}$$

$$= \frac{b_{i-1}\cdot(b_{i-1}+c-1+w_{i-1})}{(b_{i-1}+w_{i-1})\cdot(b_{i-1}+w_{i-1}+c-1)}$$

$$= \frac{b_{i-1}}{b_{i-1}+w_{i-1}}$$

$$= X_{i-1}$$

□

Lemma 1

If a sequence $X_0, X_1, \ldots$ is a martingale then,

$$\forall i : \quad E[X_i] = E[X_0]$$

Proof:
Since X_i is a martingale, by the definition we have that:

$$\begin{aligned}
\forall i : E[X_i | X_0, \ldots, X_{i-1}] &= X_{i-1} \Rightarrow \\
E\Big[E[X_i | X_0, \ldots, X_{i-1}]\Big] &= E\big[X_{i-1}\big] \Rightarrow \\
E[X_i] &= E[X_{i-1}] \Rightarrow \text{ (inductively)} \\
E[X_i] &= E[X_0], \ \forall\, i
\end{aligned}$$

□

Properties of martingales

- It is possible to construct a martingale from **any** random variable.
  - random variable $\leftrightarrow$ graph-theoretic function in random graph

  $\Rightarrow$ we can construct a martingale **for any graph-theoretic function**.
- The martingale is constructed using a generic way, as follows.

Doob Martingale

Definition 2

Consider Ω a probability sample space and $F_0, F_1, \dots$ a filter of it. Let X be any random variable that takes values in Ω.
By defining $X_i = E[X|F_i]$ the sequence $X_0, X_1, \dots$ is a Doob martingale.

Note:

A sequence $F_0, F_1, \dots$ is a filter of Ω when successive F_i consist successive refinements of it. (F_n is the most detailed refinement of Ω i.e. the sample points)

The Edge Exposure Martingale

Definition 3

Let G be random graph from $G_{n,p}$ and $f(G)$ be any graph theoretic function. Arbitrarily label the $m = \binom{n}{2}$ possible edges with the sequence $1, \ldots, m$. For $1 \leq j \leq m$, define the indicator random variable

$$I_j = \begin{cases} 1 & e_j \in G \\ 0 & otherwise \end{cases}$$

The (Doob) edge exposure martingale is defined to be the sequence of random variables $X_0, \ldots, X_m$ such that

$$X_k = E[f(G) | I_1, \ldots, I_k]$$

while $X_0 = E[f(G)]$ and $X_m = f(G)$.

The Edge Exposure Martingale - Example

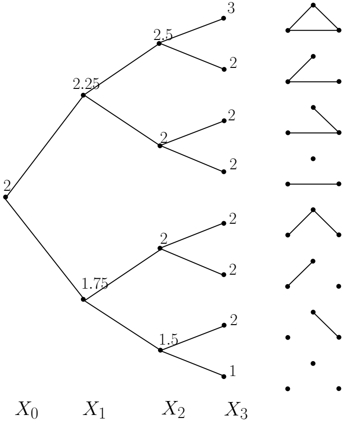

Figure: Edge exposure martingale

- $G_{n,1/2}$
- $m = n = 3$
- f = chromatic number
- The edges are exposed in the order "bottom, left, right".

The values X_k are given by tracing from the central node to leaf node.

The Edge Exposure Martingale - Example

Remarks:

- $\exists\ 2^3$ graphs (sample points), every one with probability $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$
- at time i there are i edges exposed $(i = 0, 1, 2, 3)$
- when $i = 3$ all edges are exposed and thus X_3 is the function f.
- when $i = 0$ no edge is exposed and thus $X_0 = E[f(G)]$ is constant.

$$X_0 = \frac{1}{8} \cdot (3 + 2 + 2 + 2 + 2 + 2 + 2 + 1) = \frac{1}{8} \cdot 16 = 2$$

- $\forall i : \ X_i = E[X_{i+1} | X_0, \ldots, X_i]$ since:
  - $X_2 = 2.5 = \frac{1}{2} \cdot 3 + \frac{1}{2} \cdot 2 = E[X_3 | X_0, X_1, X_2]$
  - $X_2 = 2 = \frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 2 = E[X_3 | X_0, X_1, X_2]$
  - $X_2 = 2 = \frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 2 = E[X_3 | X_0, X_1, X_2]$
  - $X_2 = 1.5 = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 2 = E[X_3 | X_0, X_1, X_2]$
  - $X_1 = 2.25 = \frac{1}{2} \cdot 2.5 + \frac{1}{2} \cdot 2 = E[X_2 | X_0, X_1]$
  - $X_1 = 1.75 = \frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 1.5 = E[X_2 | X_0, X_1]$
  - $X_0 = 2 = \frac{1}{2} \cdot 2.25 + \frac{1}{2} \cdot 1.75 = E[X_1 | X_0]$

$\Rightarrow \quad X_i$ is a martingale.

The Vertex Exposure Martingale

Definition 4

Let G be random graph from $G_{n,p}$ and $f(G)$ be any graph theoretic function. Arbitrarily label the $m = \binom{n}{2}$ possible edges with the sequence $1, \ldots, m$. Define the set $E_i \quad 1 \leq i \leq n$ as the set of all possible edges with vertices in $\{1, \ldots, i\}$. Also, $\forall j \in E_i$, define the indicator random variable

$$I_j = \begin{cases} 1 & e_j \in G \\ 0 & otherwise \end{cases}$$

Also, define the vector $\hat{I}_i = [I_1, \ldots, I_j, \ldots], \ \forall j \in E_i$.
The (Doob) vertex exposure martingale is defined to be the sequence of random variables $Y_0, \ldots, Y_n$ such that

$$Y_k = E[f(G)|\hat{I}_1, \ldots, \hat{I}_k]$$

while $Y_0 = E[f(G)]$ and $Y_n = f(G)$.

Azuma's inequality

Definition 5

Let $X_0 = 0, X_1, \dots X_m$ *be a martingale with*

$$|X_{i+1} - X_i| \leq 1$$

for all $0 \leq i < m$. *Let* $\lambda > 0$ *be arbitrary. Then*

$$\Pr\{X_m > \lambda\sqrt{m}\} < e^{-\lambda^2/2}$$

Generalization:
If $X_0 = c$ then

$$\Pr\{|X_m - c| > \lambda\sqrt{m}\} < 2e^{-\lambda^2/2}$$

Azuma's inequality importance

- Let $f(G)$ be a graph-theoretic function.
- Consider a Doob exposure martingale with
  - $X_0 = c = E[f(G)]$ and
  - X_m or Y_n = f(G)

If $|X_{i+1} - X_i| \leq 1$ then

$$\Pr\left\{\left|f(G) - E[f(G)]\right| > \lambda\sqrt{m}\right\} < 2e^{-\lambda^2/2}$$

Lipschitz condition

Definition 6

A graph-theoretic function $f(G)$ satisfies the edge (respectively vertex) Lipschitz condition iff $\forall G, G'$ that differ only in one edge (respectively vertex) it is:

$$\left| f(G) - f(G') \right| \leq 1$$

Theorem 1

If a graph-theoretic function f satisfies the edge (vertex) Lipschitz condition then the corresponding edge (vertex) exposure martingale X_i satisfies

$$|X_{i+1} - X_i| \leq 1$$

Example - Chromatic number of a random graph

Definition 7

The Chromatic number $\chi(G)$ is the least number of colors required to color the vertices of a graph so that any adjacent vertices do not have the same color.

Theorem 2

Let G be a graph in $G_{n,p}$ then

$$\forall \lambda > 0 : \quad \Pr\left\{ \left| \chi(G) - E[\chi(G)] \right| > \lambda\sqrt{n} \right\} < 2e^{-\lambda^2/2}$$

Proof of theorem 2

- Consider the Doob vertex exposure martingale $X_0, X_1 \ldots$ that corresponds to graph-theoretic function $f(G) = \chi(G)$.
- We observe that the Doob vertex exposure martingale satisfies the Lipschitz condition since the exposure of a new vertex may increase the current chromatic number $\chi(G)$ at most by 1.
- Applying theorem 1 it holds that $|X_{i+1} - X_i| \leq 1$.
- We now apply the generalized Azuma inequality with $c = X_0 = E[\chi(G)]$ and have

$$\forall \lambda > 0 : \quad \Pr\left\{ \left| \chi(G) - E[\chi(G)] \right| > \lambda\sqrt{n} \right\} < 2e^{-\lambda^2/2}$$

since $X_n = \chi(G)$

□

Example Balls and Bins

Suppose there are n balls and n bins.We are randomly throwing each ball into a bin. Define the function $L(n)$ that corresponds to the number of empty bins. Prove that

$$\forall \lambda > 0 : \quad \Pr\left\{ \left| L(n) - \frac{n}{e} \right| > \lambda\sqrt{n} \right\} < 2e^{-\lambda^2/2}$$

Proof:

- We define the indicator variable

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

- Thus, $L(n) = \sum_{i=1}^{n} l_i$ is the number of empty bins.
- $E[l_i] = 1 \cdot \Pr\{l_i = 1\} + 0 \cdot \Pr\{l_i = 0\} = \left(1 - \frac{1}{n}\right)^n \sim \frac{1}{e}$

$$\text{by L.O.E.} \quad E[L(n)] = E\left[\sum_{i=1}^{n} l_i\right] = \sum_{i=1}^{n} E[l_i] \sim n \cdot \frac{1}{e}$$

Example Balls and Bins

- Consider the Doob vertex exposure martingale $X_0, X_1 \ldots$ that corresponds to the function $L(n)$ (vertices correspond to balls).
- We observe that Doob vertex exposure martingale satisfies the Lipschitz condition since the exposure of a new vertex (i.e. the throwing a new ball in a bin) may decrease the current number of empty bins $L(n)$ at most by 1.
- Applying theorem 1 it holds that $|X_{i+1} - X_i| \leq 1$.
- We now apply generalized Azuma inequality with $c = X_0 = E[L(n)]$ and have

$$\forall \lambda > 0: \quad \Pr\left\{\left|L(n) - \frac{n}{e}\right| > \lambda\sqrt{n}\right\} < 2e^{-\lambda^2/2}$$

since $X_n = L(n)$ and $E[L(n)] \sim \frac{n}{e}$.

□