Streaming Algorithms

Meng-Tsung Tsai

03/09/2018

References

- "The Probabilistic Method", Alon and Spencer (2004)

You may find an e-copy of this book on www.lib.nctu.edu.tw

- "Chernoff Bound + Union Bound" http://www.cs.princeton.edu/courses/archive/fall09/cos521/Handouts/probabilityandcomputing.pdf
- "Martingales" https://www.cs.cmu.edu/~avrim/Randalgs11/lectures/lect0321.pdf

Monotone Graph Properties

Graph Properties

You may know some graph properties before, e.g. planarity, bipartiteness, triangle-freeness, etc.

What is a canonical representation for all graph properties?

Canonical Representation

Consider the simple undirected graphs whose node set

$$V = N = \{1, 2, ..., n\}$$

and the edge set

$$E \in P(\binom{N}{2}) \text{ where } P(\binom{N}{2}) \text{ is the power set of } \binom{N}{2},$$

i.e. the family of all subsets of $\binom{N}{2}$.

Then one can define each graph property as a subset $Q \subseteq P(\binom{N}{2})$ so that for every graph G = (V, E) that satisfies the graph property, we have $E \in Q$.

How many "different" graph properties for n-node graphs?

Canonical Representation

Consider the simple undirected graphs whose node set

$$V = N = \{1, 2, ..., n\}$$

and the edge set

$$E \in P(\binom{N}{2}) \text{ where } P(\binom{N}{2}) \text{ is the power set of } \binom{N}{2},$$

i.e. the family of all subsets of $\binom{N}{2}$.

Then one can define each graph property as a subset $Q \subseteq P(\binom{N}{2})$ so that for every graph G = (V, E) that satisfies the graph property, we have $E \in Q$.

How many "different" graph properties for n-node graphs?

$2^{\binom{n}{2}}$

Monotone Increasing Properties

We say Q is a monotone increasing graph property

$$\text{if } E \subseteq E' \text{ and } E \in Q \text{ then } E' \in Q.$$

Is there a graph property monotone increasing?

Connectivity, Hamiltonian, etc.

Monotone Decreasing Properties

We say Q is a monotone decreasing graph property

$$\text{if } E' \subseteq E \text{ and } E \in Q \text{ then } E' \in Q.$$

Is there a graph property monotone decreasing?

Bipartiteness, Triangle-freeness, etc.

Systematic Study

Some graph properties have similar behaviors, and we may study the graph properties in a certain class by a unified way.

Theorem 1. Let Q_1, Q_2, Q_3 and Q_4 be graph properties. Q_1 and Q_2 are monotone increasing. Q_3 and Q_4 are monotone decreasing. Let G be a random graph sampled from G(n, p), i.e. an n-node graph so that each edge is included with probability p independently. Then,

$$(1)\ \Pr[G \in Q_1 \cap Q_2] \geq \Pr[G \in Q_1]\Pr[G \in Q_2]$$

$$(2)\ \Pr[G \in Q_3 \cap Q_4] \geq \Pr[G \in Q_3]\Pr[G \in Q_4]$$

$$(3)\ \Pr[G \in Q_1 \cap Q_3] \leq \Pr[G \in Q_1]\Pr[G \in Q_3]$$

Proof

A proof can be found on Pages 82-88 in the reference book.

Example Application

Let $G \sim G(2k, 1/2)$ and let $\Delta(G)$ denote the maximum degree of G. We have:

$$\Pr[\Delta(G) < k] \geq 1/4^k.$$

Example Application

Let $G \sim G(2k, 1/2)$ and let $\Delta(G)$ denote the maximum degree of G. We have:

$$\Pr[\Delta(G) < k] \geq 1/4^k.$$

Proof. Let x be a certain node in G. Then $\Pr[\deg(x) < k] = 1/2$. (Why?) Since $[\Delta(G) < k] = [\wedge_i \deg(x_i) < k]$ and each $[\deg(x_i) < k]$ is a monotone decreasing property,

$$\Pr[\Delta(G) < k] \geq \prod_i \Pr[\deg(x_i) < k] = 1/4^k.$$

Systematic Study

Some graph properties have similar behaviors, and we may study the graph properties in a certain class by a unified way.

Theorem 2. For each nontrivial hereditary property Q, given a graph G, deleting the minimum number of nodes in G so that Q is satisfied is NP-hard. (aka the node-deletion problem for property Q)

Nontrivial: Q is infinite.

Hereditary: If G satisfies Q, then every node-induced subgraph H of G satisfies Q.

differences between node-induced subgraphs and subgraphs

Exercise 1

Prove or disprove the node-deletion problem for connectivity can be solved in polynomial time. You may assume that $P \neq NP$.

Chernoff Bounds

Chernoff Bounds

Let $X_1, X_2, ..., X_n$ be mutually independent random variables (unnecessary to have identical distributions), and $0 \leq X_i < 1$ for each i. Let $X = \sum_i X_i$ and $\boldsymbol{\mu} = E[X] = \sum_i E[X_i]$.

Chernoff Bounds state that for any $\varepsilon \geq 0$,

$$Pr[X \geq (1+\varepsilon)\boldsymbol{\mu}] \leq \exp[-\varepsilon^2\boldsymbol{\mu}/(2+\varepsilon)] \text{ and}$$
$$Pr[X \leq (1-\varepsilon)\boldsymbol{\mu}] \leq \exp[-\varepsilon^2\boldsymbol{\mu}/2].$$

Proof can be found in the second reference.

Example Application

The bin-ball problem has the following setting. We have n bins and m balls. Each ball is thrown i.i.d. uniformly into one of the n bins.

If $m = C\, n \log n$ for any constant $C > 1$,
then $\Pr[\text{no bin is empty}] = 1/n^{\Omega(1)}$.

$\Pr[\text{a certain bin is empty}] = (1-1/n)^m \sim 1/n^C$.

By Union bound, $\Pr[\text{no bin is empty}] \leq n/n^C = 1/n^{\Omega(1)}$.

Example Application

The bin-ball problem has the following setting. We have n bins and m balls. Each ball is thrown i.i.d. uniformly into one of the n bins.

If $m = n \log n$,
then $\Pr[\text{max bin has } O(\log n) \text{ balls}] = 1-1/n^{\Omega(1)}$.

$\Pr[\text{a certain bin has } O(\log n) \text{ balls}] \sim 1/n^C$ for any constant C.

By Union bound, $\Pr[\text{max bin has } O(\log n) \text{ balls}] \leq n/n^C = 1/n^{\Omega(1)}$.

More Applications

Refer to the second reference for more applications of Chernoff bounds + Union bounds.

Martingale

Discrete-time Martingale

Let $X_0, X_2, \ldots, X_n$ be a sequence of random variables that satisfy:

(1) $E[|X_n|] < \infty$

(2) $E[X_{i+1} \mid X_i, X_{i-1}, \ldots, X_0] = X_i$ for each $1 \leq i < n$.

A Simple Martingale

We flip n coins uniformly i.i.d. Let $X_i \in \{+1, -1\}$ so that $X_i = +1$ iff the i-th flipping has head up, and $Y_i = \sum_{1 \leq j \leq i} X_i$.

Clearly, $E[|Y_n|] \leq n < \infty$.

$E[Y_{i+1} \mid Y_i, Y_{i-1}, \ldots, Y_0] = Y_i$ for each $0 \leq i < n$ because

$$E[Y_{i+1} \mid Y_i, Y_{i-1}, \ldots, Y_0] = Y_i + E[X_i] = Y_i.$$

An Edge-exposure Martingale

Let $G \sim G(n, p)$ and let $\chi(G)$ be the chromatic number of G, i.e. the smallest number of colors needed to color nodes in G so that no two adjacent nodes share the same color.

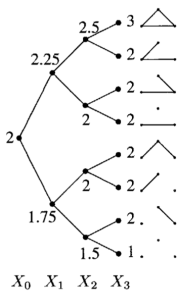
Example. $n = 3$, $p = 1/2$.

$X_0 = E[\chi(G)]$.

$X_1 = E[\chi(G) \mid E_b]$ (reveal the bottom edge)

$X_2 = E[\chi(G) \mid E_\ell, E_b]$ (reveal the left edge)

$X_3 = E[\chi(G) \mid E_r, E_\ell, E_b]$ (reveal the right edge)



This is not a formal proof.

More Martingales

You can found more martingales in the 3rd reference.

Azuma's Inequality

Let $X_0 = c, X_1, \ldots, X_m$ be a martingale and $|X_{i+1} - X_i| \leq 1$ for each i in [0, m-1]. Then

$$\Pr[|X_m - c| > \lambda m^{1/2}] < 2\exp(-\lambda^2/2).$$

Proof can be found on Page 95 in the refernce book.

See a more general statement in Wikipedia https://en.wikipedia.org/wiki/Azuma%27s_inequality

Chromatic Number

Let $G \sim G(n, p)$ and let $\chi(G)$ be the chromatic number of G. Then, one can define the node-exposure martingale $X_0 = c, X_1, \ldots, X_n$ accordingly, i.e. in the i-th round revealing all (still) random edges in G that has end-points in $\{1, 2, \ldots, i\}$.

$$\Pr[|X_m - c| > \lambda (n-1)^{1/2}] < 2\exp(-\lambda^2/2).$$

$|X_{i+1} - X_i| \leq 1$ because adding a new node to a graph cannot increase the chromatic number by more than one, so does the expected value.

Bin-Ball Problem

Let L(n, n) denote # of empty bins in the bin-ball problem with m = n.

Thus, $E[L(n, n)] = n(1-1/n)^n \sim n/e$.

L(n, n) is a martingale if one reveal the bin where the i-th ball lands in one by one. (See the 3rd reference)

$$\Pr[|L(n, n) - n/e| > \lambda n^{1/2}] < 2\exp(-\lambda^2/2).$$