Randomness in Computing

CS 537

LECTURE 16

Last time before review

- Poisson approximation

Today

- Discuss midterm
- Review Poisson approximation
- Application: max load
- Application: Coupon Collector
- Random graphs

CS 537

Review question

Let X and Y be random variables that take values in $\mathbb{N}$.

Which of the following are NOT well defined?

A. $\Pr[X]$

B. $\Pr[X = Y]$

C. $\Pr[X \cap Y]$

D. $\Pr[X \text{ is a geometric RV with parameter } p]$

E. $\mathbb{E}[\mathbb{E}[X]]$

F. $\mathbb{E}[X = 0]$

Make sure your objects are of the right type!

Review: Poisson random variables

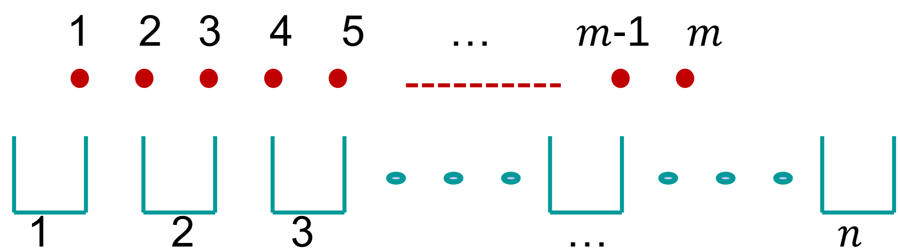
- A Poisson random variable with parameter μ is given by the following distribution on $j = 0,1,2, \ldots$

$$\Pr[X = j] = \frac{e^{-\mu}\mu^j}{j!}$$

- $\mathbb{E}[X] = \mathrm{var}[X] = \mu$
- Sum of two Poisson RVs is a Poisson RV
- A Poisson RV satisfies Chernoff-type bounds.
- Poisson distribution is a limit of binomial distribution

Review: Poisson approximation

- The Balls-and-Bins model has dependences.
- The Poisson approximation gets rid of dependencies.



- m balls into n bins u.i.r.

  For $i \in [n]$, let

  (real world) $X_i^{(m)} = \#$ of balls in bin i

  (Poisson world) $Y_i^{(m)} \sim Poisson(\mu)$, where $\mu = \frac{m}{n}$ and

  $Y_i^{(m)}$ are mutually independent.

- If we condition the Poisson distribution on producing exactly k balls, then it's the same as the distribution resulting from throwing k balls into n bins.

Review: Approximating a function of the loads of the bins

Poisson Approximation Theorem

Let $f(x_1, \dots, x_n) \geq 0$ for all $x_1, \dots, x_n \in \{0,1,2,\dots\}$. Then

$$\mathbb{E}\left[f\left(X_1^{(m)}, \dots, X_n^{(m)}\right)\right] \leq \boldsymbol{e\sqrt{m}} \cdot \mathbb{E}\left[f\left(Y_1^{(m)}, \dots, Y_n^{(m)}\right)\right].$$

- Poisson case: # of balls in each bin is independent Poisson$\left(\frac{m}{n}\right)$
- Corollary. Any event that has probability p in the Poisson case has probability $\leq p \cdot \boldsymbol{e\sqrt{m}}$ in the exact case.

  Proof: Let X be the indicator for that event.

  Then $\mathbb{E}[X]$ is the probability that event occurs.
- Improvements to Theorem and Corollary

  If $\mathbb{E}\left[f\left(X_1^{(m)}, \dots, X_n^{(m)}\right)\right]$ is monotonically nonincreasing (or nondecreasing) in m, then $\boldsymbol{e\sqrt{m}}$ can be changed to 2.

Application: Max Load

***n* balls into *n* bins**

- Before (by Chernoff): $\Pr\left[MaxLoad > \frac{3\ln n}{\ln\ln n}\right] \le \frac{1}{n}$ for s.l. n
- Theorem. $\Pr\left[MaxLoad < \frac{\ln n}{\ln\ln n}\right] \le \frac{1}{n}$ for s.l. n

Proof: Let $M = \frac{\ln n}{\ln\ln n}$

Application: Coupon Collector

X = # of coupons observed before obtaining 1 of each of n types

- Before: $\mathbb{E}[\mathrm{X}] =$ and $\Pr[X > n\ln n + cn] \le e^{-c}\ \forall c > 0$

  Review: Pr[not obtaining coupon i in $n \ln n + cn$ steps] is

- Theorem. $\Pr[X > n\ln n + cn] \le 2(1 - e^{-1.5 \cdot e^{-c}})\ \ \forall c > 0$, s.l. n
- MU: $\lim_{n\to\infty} \Pr[X > n\ln n + cn] = 1 - e^{-e^{-c}}\ \ \forall c > 0$

Application: Coupon Collector

X = # of coupons observed before obtaining 1 of each of n types

- Theorem. $\Pr[X > n \ln n + cn] \leq 2(1 - e^{-1.5 \cdot e^{-c}})\ \ \forall c > 0$, s.l. n

Proof: Balls and bins view

n bins

- X = # balls thrown before all bins nonempty
- Consider $m = n(\ln n + c)$ balls.
- Let B = event that there is an empty bin.

$$\Pr[X > n \ln n + cn] = \Pr[\mathrm{B}]$$

- Idea: Use Poisson approximation:

balls in each bin is Poisson(μ) with $\mu = \ln n + c$

- E_i = event that bin i is empty. Then
- $\Pr[E_i] = e^{-\mu} = e^{-(\ln n + c)} = \frac{e^{-c}}{n}$

Application: Coupon Collector

Theorem. $\Pr[X > n \ln n + cn] \leq 2(1 - e^{-1.5 \cdot e^{-c}})\ \ \forall c > 0$, s.l. n

Proof: n bins and $m = n(\ln n + c)$ balls.

- Let B = event that there is an empty bin: $\Pr[X > n \ln n + cn] = \Pr[\mathrm{B}]$
- Poisson approximation: # balls in each bin is Poisson(μ), $\mu = \ln n + c$
- E_i = event that bin i is empty. Then $\Pr[E_i] = \frac{e^{-c}}{n}$

$\boldsymbol{0 \leq}$ ***for*** $\boldsymbol{x \leq 1/2}$

$\boldsymbol{e^{-1.5x} = e^{-x-0.5\,x} \leq}$

$\boldsymbol{e^{-x-x^2} \leq 1 - x \leq e^{-x}}$

Random graph models

$\boldsymbol{G_{n,p}}$: n vertices, each of $\binom{n}{2}$ potential edges is present w.p. p

$\boldsymbol{G_{n,M}}$: n vertices, a uniformly random M-edge graph

Question

Suppose $G \sim G_{n,p}$

What is the expected number of edges in G?

A. p

B. n/p

C. pn

D. pn^2

E. $pn(n-1)/2$

F. None of the above

Question

The distribution $G_{n,p}$ conditioned on producing a graph with M edges is

A. $G_{n,M}$

B. $G_{n,pn}$

C. $G_{n,p\binom{n}{2}}$

D. None of the above

Random graph models

$\boldsymbol{G_{n,p}}$: n vertices, each of $\binom{n}{2}$ potential edges is present w.p. p

$\boldsymbol{G_{n,M}}$: n vertices, a uniformly random M-edge graph

- When $p = \frac{M}{\binom{n}{2}}$, the number of edges in a graph $\sim G_{n,p}$ is concentrated around M.
- Conditioning claim. $G_{n,p}$ conditioned on having M edges $\Leftrightarrow G_{n,M}$

Graph properties

Graph property: whether a graph has it doesn't depend on how its vertices are labeled (it holds for $G \Leftrightarrow$ it holds for all graphs isomorphic to G).

A graph property is monotone *increasing* if:

for every graph G that has the property, every graph G' obtained by *adding* edge(s) to G also has the property.

Examples: connected, has a component with $\geq k$ vertices, has a cycle, …

A graph property is monotone *decreasing* if:

for every graph G that has the property, every graph G' obtained by *deleting* edge(s) from G also has the property.

Examples: acyclic, bipartite, is a forest…

Not monotone: **e.g.**, is a tree.

Relating $G_{n,M}$ and $G_{n,p}$

Theorem

Consider a monotone increasing graph property.

Let $\boldsymbol{P(n,M)}$ be the probability it holds for a graph $\sim G_{n,M}$.

Let $\tilde{P}(n,p)$ be the probability it holds for a graph $\sim G_{n,p}$.

Fix a constant $\epsilon \in (0,1)$. Let $p^- = (1-\epsilon)\cdot\frac{M}{\binom{n}{2}}$ and $p^+ = (1+\epsilon)\cdot\frac{M}{\binom{n}{2}}$. Then

$$\tilde{P}(n,p^-) - e^{-\Omega(M)} \leq \boldsymbol{P(n,M)} \leq \tilde{P}(n,p^+) + e^{-\Omega(M)}$$

Proof: Let R.V. X = # of edges in a graph chosen $\sim G_{n,p^-}$.

$$\tilde{P}(n,p^-) = \sum_{k=0}^{\binom{n}{2}} P(n,k)\cdot \Pr[X=k]$$

$$= \sum_{k\leq M} P(n,k)\cdot \Pr[X=k] + \sum_{k>M} \overset{\leq 1}{P(n,k)}\cdot \Pr[X=k]$$

$$\leq \boldsymbol{P(n,M)}\cdot \overset{\leq 1}{\Pr[X\leq M]} + \Pr[X>M]$$

$$\leq \boldsymbol{P(n,M)} + \Pr[X>M]$$

By Law of Total Probability + Conditioning Claim

Mon. increasing property $\Rightarrow$ $\boldsymbol{P(n,k) \leq P(n,M)\ \forall k \leq M}$

Relating $G_{n,M}$ and $G_{n,p}$

Theorem

Consider a monotone increasing graph property.

Let $\boldsymbol{P(n,M)}$ be the probability it holds for a graph $\sim G_{n,M}$.

Let $\tilde{P}(n,p)$ be the probability it holds for a graph $\sim G_{n,p}$.

Fix a constant $\epsilon \in (0,1)$. Let $p^- = (1-\epsilon)\cdot\frac{M}{\binom{n}{2}}$ and $p^+ = (1+\epsilon)\cdot\frac{M}{\binom{n}{2}}$. Then

$$\tilde{P}(n,p^-) - e^{-\Omega(M)} \le \boldsymbol{P(n,M)} \le \tilde{P}(n,p^+) + e^{-\Omega(M)}$$

Proof: Let R.V. X = # of edges in a graph chosen $\sim G_{n,p^-}$.

$\tilde{P}(n,p^-) \le \boldsymbol{P(n,M)} + \Pr[X > M]$

Apply Chernoff:

$\mu = \mathbb{E}[X] = (1-\epsilon)M$

since $\epsilon \in (0,1)$

$$\Pr[X > M] = \Pr\left[X > \underbrace{\frac{1}{1-\epsilon}}_{\ge 1+\epsilon}\cdot\mu\right] \le \Pr[X > (1+\epsilon)\cdot\mu]$$

$$\le e^{-\frac{\epsilon^2\mu}{3}} = e^{-\frac{\epsilon^2(1-\epsilon)M}{3}}.$$

Relating $G_{n,M}$ and $G_{n,p}$

Theorem

Consider a monotone increasing graph property.

Let $\boldsymbol{P(n,M)}$ be the probability it holds for a graph $\sim G_{n,M}$.

Let $\tilde{P}(n,p)$ be the probability it holds for a graph $\sim G_{n,p}$.

Fix a constant $\epsilon \in (0,1)$. Let $p^- = (1-\epsilon)\cdot\frac{M}{\binom{n}{2}}$ and $p^+ = (1+\epsilon)\cdot\frac{M}{\binom{n}{2}}$. Then

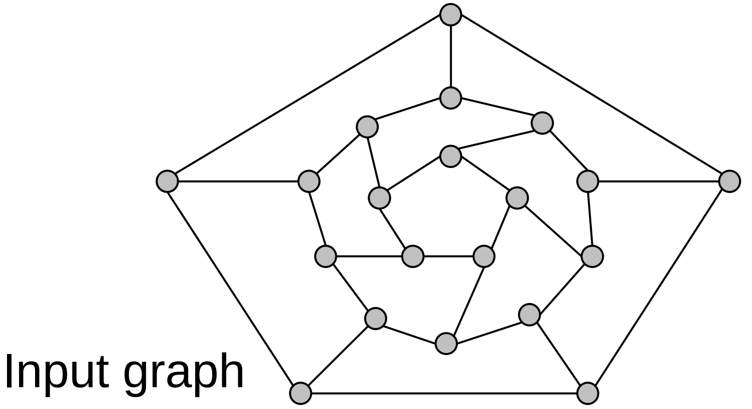
$$\tilde{P}(n,p^-) - e^{-\Omega(M)} \le \boldsymbol{P(n,M)} \le \tilde{P}(n,p^+) + e^{-\Omega(M)}$$

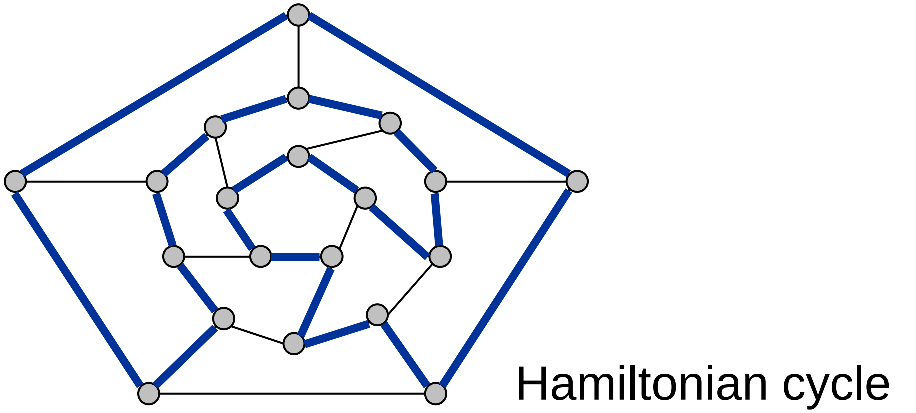
Proof: The proof for the second inequality is similar.

A similar theorem holds for monotone decreasing properties.

Finding Hamiltonian cycles

A Hamiltonian cycle is a cycle that visits each vertex exactly once.





- Finding Hamiltonian cycles is NP-hard.

Question

The property of having a Hamiltonian cycle is

A. Monotone increasing

B. Monotone decreasing

C. Not monotone

CS 537

Finding Hamiltonian cycles in random graphs

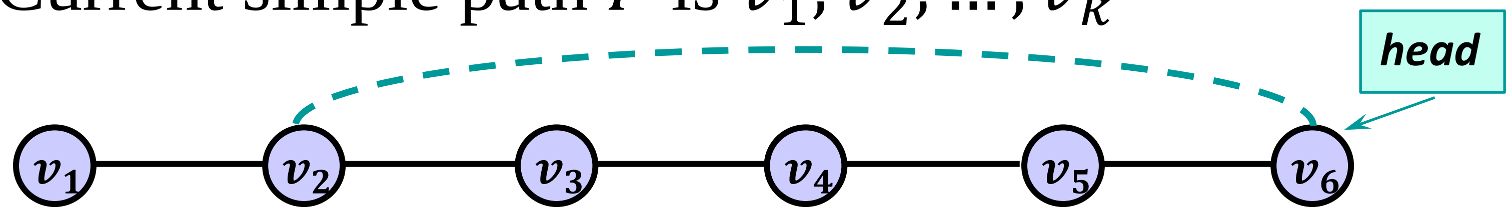
Main Theorem

Suppose $p \geq \frac{40 \ln n}{n}$. There is a polynomial time randomized algorithm that, given p and $G \sim G_{n,p}$, finds a Hamiltonian cycle in G with probability $1 - O\left(\frac{1}{n}\right)$.

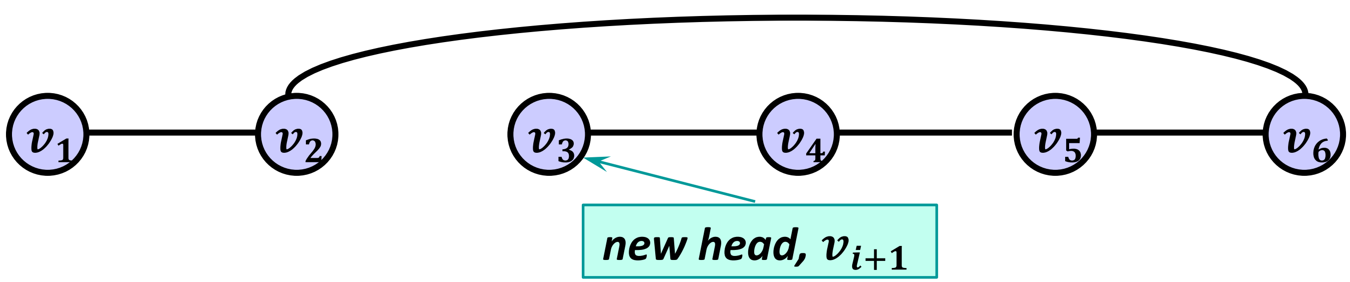
Corollary: Hamiltonian cycle exists in $G \sim G_{n,p}$ w. p. $1 - O\left(\frac{1}{n}\right)$.

Rotation operation

- Current simple path P is $v_1, v_2, \ldots, v_k$



- Rotation of P with an edge (v_k, v_i) in G, $i \in [k-1]$



  – If $i = k - 1$, no change.

  – If $k = n$ and $i = 1$,
  rotation edge (v_k, v_i) closes Hamiltonian path.