Final Preparation for Probabilistic Method and Random Graphs

Rui Zhang

ICT, CAS
zhangrui2016@ict.ac.cn

Overview

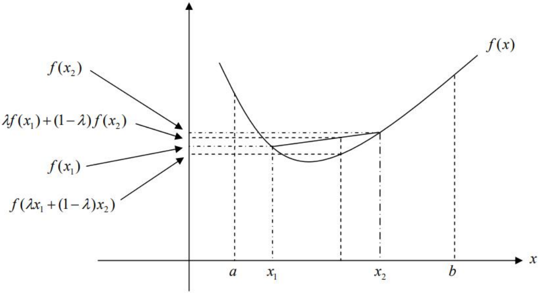
Overview

- Linearity of Expectation

$$\mathbf{E}\sum X_i = \sum \mathbf{E}X_i$$

- Jensen equality: for convex f:

$$f(\mathbf{E}X) \leq \mathbf{E}f(X)$$



Union bound $\mathbf{P}(\cup E_i) \leq \sum \mathbf{P}(E_i)$

- given a set $\mathscr{A}$

$$\mathbf{P}(\max \mathscr{A} > c) = \mathbf{P}(\exists a \in \mathscr{A} : a > c) \leq \sum_{a \in \mathscr{A}} \mathbf{P}(a > c)$$

Chernoff Bound Technique

$$e^x \geq 1 + x$$

1. Prove the following extensions of the Chernoff bound. Let $X = \sum_{i=1}^n X_i$, where the X_i's are independent Poisson trials. Let $\mu = \mathbb{E}[X]$. Choose any μ_L and μ_H such that $\mu_L \leq \mu \leq \mu_H$. Then, for any $\delta > 0$, $\Pr(X \geq (1+\delta)\mu_H) \leq (\frac{e^\delta}{(1+\delta)^{(1+\delta)}})^{\mu_H}$.

$$\mathbf{P}\left(X \geq (1+\delta)\mu_H\right) = \mathbf{P}\left(e^{\lambda X} \geq e^{\lambda(1+\delta)\mu_H}\right) \leq \frac{\mathbf{E}[e^{\lambda X}]}{e^{\lambda(1+\delta)\mu_H}}$$

$$= \frac{\mathbf{E}[e^{\lambda \sum_{i=1}^n X_i}]}{e^{\lambda(1+\delta)\mu_H}} = \frac{\prod_{i=1}^n \mathbf{E}[e^{\lambda X_i}]}{e^{\lambda(1+\delta)\mu_H}} = \frac{e^{\sum_{i=1}^n (p_i e^\lambda + (1-p_i))}}{e^{\lambda(1+\delta)\mu_H}} \leq \frac{e^{\sum_{i=1}^n p_i (e^\lambda - 1)}}{e^{\lambda(1+\delta)\mu_H}}$$

$$= \frac{e^{\mu(e^\lambda - 1)}}{e^{\lambda(1+\delta)\mu_H}} \leq \frac{e^{\mu_H(e^\lambda - 1)}}{e^{\lambda(1+\delta)\mu_H}} = \left(\frac{e^{(e^\lambda - 1)}}{e^{\lambda(1+\delta)}}\right)^{\mu_H}$$

Bins and Balls Model

m balls, n bins

$$\mathbf{P}(\text{the number of balls in bin i } \geq k) \leq \binom{m}{k}\frac{1}{n^k}$$

$$\mathbf{P}(\text{max load} \geq k) \leq n\binom{m}{k}\frac{1}{n^k}$$

$$\mathbf{P}(X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n) = \frac{m!}{k_1!k_2!\ldots k_n!n^m}$$

First Moment method and De-randomization

$$\mathbf{P}(X \geq \mathbf{E}X) > 0$$

$$\begin{aligned}\mathbf{E}X = \sum_i i\mathbf{P}(X=i) = & \sum_{i \geq c} i\mathbf{P}(X=i) + \sum_{i<c} i\mathbf{P}(X=i) \\ \leq & (\max X)\mathbf{P}(X \geq c) + c(1-\mathbf{P}(X \geq c))\end{aligned}$$

De-randomization

$\mathbf{X} = (x_1, \ldots, x_n)$

$$x_k = \arg\min_{v_k} \mathbf{E}[f(\mathbf{X}) \mid x_1 = v_1, \ldots, x_{k-1} = v_{k-1}, x_k = v_k]$$

Second Moment Method

- Markov $X \in [0, \infty)$

$$\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}X}{a}$$

- Chebyshev

$$\mathbf{P}(|X - \mathbf{E}X| \geq t) \geq \frac{\mathrm{Var}\, X}{t^2}$$

$$\mathbf{P}(X = 0) \leq \mathbf{P}(|X - \mathbf{E}X| \geq \mathbf{E}X) \leq \frac{\mathrm{Var} X}{(\mathbf{E}X)^2}$$

LLL

d: max degree

$$4pd < 1$$

Given real numbers: $x_1, \ldots, x_n \in [0, 1)$
$\Gamma(i)$ is the set of neighbors of vertex i

$$\forall i \in [n] : \mathbf{P}(i) \leq x_i \prod_{j \in \Gamma(i)} (1 - x_j)$$

Overview

1. Prove that any memoryless distribution on positive integers is a geometric distribution.

Memoryless implies

$$\forall m,n \in \mathbb{Z}^+ : \mathbf{P}(X > m+n \mid X > m) = \mathbf{P}(X > n)$$

or

$$\forall m,n \in \mathbb{Z}^+ : \frac{\mathbf{P}(X > m+n \cap X > m)}{\mathbf{P}(X > m)} = \frac{\mathbf{P}(X > m+n)}{\mathbf{P}(X > m)} = \mathbf{P}(X > n)$$

or

$$\forall m,n \in \mathbb{Z}^+ : \mathbf{P}(X > m+n) = \mathbf{P}(X > m)\mathbf{P}(X > n)$$

2. Assume that on an island, each couple give birth to babies until they have a female baby and a male baby. Suppose that a baby will be male or female with probability 0.5. On average how many male/female babies does a couple have? What if each couple refuses to have more than 5 babies?

Coupon collector's problem

$$\mathbf{P}(\text{the probability of collecting i-th new coupon}) = p_i = 1 - \frac{i-1}{n}$$

$$\begin{aligned}
\mathbf{E}(T) &= \mathbf{E}(t_1) + \mathbf{E}(t_2) + \cdots + \mathbf{E}(t_n) \\
&= \frac{1}{p_1} + \frac{1}{p_2} + \cdots + \frac{1}{p_n} \\
&= \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{1} \\
&= n \cdot \left(\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n}\right) \\
&= n \cdot H_n
\end{aligned}$$

H_n Harmonic series

2. Let $X_1, ...X_n$ be independent Poisson trials such that $\Pr(X_i = 1) = p_i$ and let $a_1, ...a_n$ be real numbers in $[0, 1]$. let $X = \sum_{i=1}^{n} a_i X_i$ and $\mu = \mathbb{E}[X]$. Then the following Chernoff bound holds: for any $\delta > 0$, $\Pr(X \geq (1+\delta)\mu) \leq (\frac{e^\delta}{(1+\delta)^{(1+\delta)}})^\mu$. Also prove a similar bound for the probability $\Pr(X \leq (1-\delta)\mu)$ for any $0 < \delta < 1$.

we need the following

$$\mathbf{E}[e^{\lambda a_i X_i}] = p_i e^{\lambda a_i} + 1 - p_i = 1 + p_i\left(e^{\lambda a_i} - 1\right) \leq \underline{e^{p_i(e^{\lambda a_i}-1)} \leq e^{p_i a_i(e^{\lambda}-1)}}$$

or

$$e^{\lambda a_i} - 1 \leq a_i\left(e^{\lambda} - 1\right)$$

or

$$\frac{e^{\lambda a_i} - 1}{a_i} \leq \frac{e^{\lambda * 1} - 1}{1}$$

this is slope of line through $\left(x, e^{\lambda x}\right)$ and $(0, 1)$, which is obvious via plot of function $e^{\lambda x}$

A function f is said to be convex if it holds that $f(\lambda x_1+(1-\lambda)x_2) \leq \lambda f(x_1)+(1-\lambda)f(x_2)$ for any x_1, x_2 and $0 \leq \lambda \leq 1$.

- Let Z be a random variable that takes on a finite set of values in $[0,1]$, and let $p = \mathbb{E}[Z]$. Define the Bernoulli random variable X by $\Pr(X = 1) = p$ and $\Pr(X = 0) = 1 - p$. Show that $\mathbb{E}[f(Z)] \leq \mathbb{E}[f(X)]$ for any convex function f. (Hint: by induction on the number of values that Z takes on.)

$$\mathbf{E}f(Z) = \sum_i p_i f(z_i) = \sum p_i f(z_i * 1 + (1 - z_i) * 0)$$

$$\leq \sum_i p_i z_i f(1) + p_i(1 - z_i) f(0) = pf(1) + (1 - p) f(0) = \mathbf{E}f(X)$$

1. Suppose that balls are thrown randomly into n bins. Show, for some constant c_1, that if there are $c_1\sqrt{n}$ balls then the probability that no two land in the same bin is at most $1/e$. Similarly, show for some constant c_2 (and sufficiently large n) that, if there are $c_2\sqrt{n}$ balls, then the probability that no two land in the same bin is at least $1/2$. Make these constants as close to optimum as possible. Hint: you may need the fact that $e^{-x} \geq 1 - x$ and $e^{-x-x^2} \leq 1 - x$ for $x \leq 1/2$.

the exact probability is

$$\mathbf{P}(n\ bins\ m\ balls\ max\ load = 1) = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\left(1 - \frac{m-1}{n}\right)$$

we need $\mathbf{P}(n\ bins\ m\ balls\ max\ load = 1) \leq \frac{1}{e}$ or

$$\mathbf{P}(n\ bins\ m\ balls\ max\ load = 1) \leq e^{-\frac{1}{n}} e^{-\frac{2}{n}} \ldots e^{-\frac{m-1}{n}} = e^{-\frac{m(m-1)}{2n}} \leq \frac{1}{e}$$

we need $\mathbf{P}(n\ bins\ m\ balls\ max\ load = 1) \geq \frac{1}{2}$ or

$$\mathbf{P}(n\ bins\ m\ balls\ max\ load = 1) \geq e^{-\frac{1}{n}-\frac{1}{n^2}} e^{-\frac{2}{n}-\frac{2^2}{n^2}} \ldots e^{-\frac{m-1}{n}-\frac{(m-1)^2}{n^2}} = e^{-\frac{m(m-1)}{2n}-\frac{(m-1)m(2m}{6n^2}}$$

1. (**Bonus score 5 points**) Prove the Poisson convergence theorem with weak dependence. Namely, for each n, suppose there are random variables $X_1^n, ..., X_n^n \in \{0,1\}$ such that

- $\lim_{n\to\infty} \mathbb{E}[Y_n] = \lambda$ where $Y_n = \sum_{i=1}^n X_i^n$, and
- For any k, $\lim_{n\to\infty} \sum_{1\le i_1<...<i_k\le n} \Pr\left(X_{i_1}^n = X_{i_2}^n = ... = X_{i_r}^n = 1\right) = \lambda^k/k!$

Then $\lim_{n\to\infty} Y_n \sim Poi(\lambda)$, i.e. $\lim_{n\to\infty} \Pr(Y_n = k) = e^{-\lambda}\lambda^k/k!$ for any integer $k \ge 0$. (Hint: you may need Bonferroni inequalities)

Brun's sieve, omit

The system evolves over rounds. Every round, balls are thrown independently and uniformly at random into n bins. Any ball that lands in a bin by itself is served and removed from consideration. The remaining balls are thrown again in the next round. We begin with n balls in the first round, and we will finish when every ball is served.

- If there are b balls at the start of a round, what is the expected number of balls at the start of the next round?
- Suppose that every round the number of balls served was exactly the expected number of balls to be served. Show that all the balls would be served in $O(\ln \ln n)$ rounds. (Hint: If x_j is the expected number of balls left after j rounds, show and use that $x_{j+1} \leq x_j^2/n$.)

probability of a bin with load 1

$$\mathbf{P}(X_i = 1) = \binom{b}{1}\frac{1}{n}\left(1 - \frac{1}{n}\right)^{b-1}$$

the expected balls will be served

$$\mathbf{E}X = n\mathbf{P}(X_i = 1) = b\left(1 - \frac{1}{n}\right)^{b-1}$$

thus, expected number of balls at the start of the next round
$b - b\left(1 - \frac{1}{n}\right)^{b-1}$

$$x_{j+1} = x_j - x_j\left(1 - \frac{1}{n}\right)^{x_j-1} = x_j\left[1 - \left(1 - \frac{1}{n}\right)^{x_j-1}\right]$$

consider $f(x) = \left(1 - \frac{1}{n}\right)^{x+1} - \left(1 - \frac{x}{n}\right)$, we can get $x_{j+1} \le \frac{x_j^2}{n}$ or
$\ln x_{j+1} \le 2\ln x_j - \ln n$

1. We mentioned a probabilistic proof of Turán theorem in the lecture notes. Recall the random process generating an independent set S. Let p be the probability that the independent set S has size at least $\frac{|V|}{D+1}$. Show that $p \geq \frac{1}{2D|V|^2}$.

we need

$$p = \mathbf{P}\left(|S| \geq \frac{|V|}{D+1}\right) \geq \frac{1}{2D|V|^2}$$

since

$$\mathbf{E}|S| \geq \frac{|V|}{D+1}$$

$$\mathbf{E}|S| = \sum_{|S| \geq \frac{|V|}{D+1}} |S|\mathbf{P}(|S|) + \sum_{|S < \frac{|V|}{D+1}} |S|\mathbf{P}(|S|) \leq |V|p + \left(\frac{|V|}{D+1} - 1\right)(1-p)$$

thus

$$|V|p + \left(\frac{|V|}{D+1} - 1\right)(1-p) \geq \frac{|V|}{D+1}$$

thus

$$p\left(\frac{D|V|}{D+1} + 1\right) \geq 1$$

thus

$$p \geq \frac{D+1}{D+1+D|V|} > \frac{1}{2D|V|^2}$$

3. Suppose H is a hypergraph where each edge has r vertices and meets at most d other edges. Assume that $d \leq 2^{r-3}$. Prove that H is 2-colorable, i.e. one can color the vertices in red or blue so that no monochromatic edges exist.

$$p = \frac{2}{2^r}$$

$$d \leq 2^{r-3}$$

$$4pd = 4\frac{2}{2^r}d \leq 4\frac{2}{2^r}2^{r-3} = 1$$

Use the Lovász Local Lemma to show that, if

$$4\binom{k}{2}\binom{n}{k-2}2^{1-\binom{k}{2}} \le 1,$$

then it is possible to color the edges of K_n with two colors so that it has no monochromatic K_k subgraphs. Note that this is better than the result obtained by counting.

$$p = \frac{2}{2^{\binom{k}{2}}}$$

$$d \le \binom{k}{2}\binom{n}{k-2}$$

$$4pd = 4\binom{k}{2}\binom{n}{k-2}2^{1-\binom{k}{2}}$$

1. Consider a graph in $G_{n,p}$ with $p = c\frac{\ln n}{n}$. Use the second moment method to prove that if $c < 1$ then, for any constant $\epsilon > 0$ and for n sufficiently large, the graph has isolated vertices with probability at least $1 - \epsilon$.

let X_i indicate if a vertex $v_i \in V$ is isolated

$X = \sum_{i=i}^{n} X_i$ is total number of isolated vertices

$$\mathbf{E}X = \mathbf{E}\sum_{i=i}^{n} X_i = \mathbf{E}\sum_{i=i}^{n}(1-p)^{n-1} = n(1-p)^{n-1}$$

$$
\begin{aligned}
\mathrm{Var}X &= \mathbf{E}X^2 - (\mathbf{E}X)^2 \\
&= \mathbf{E}(\sum_{i=i}^{n} X_i)^2 - (\mathbf{E}X)^2 \\
&= \sum_{i\neq j} \mathbf{E}(X_i X_j) + \sum_{i=i}^{n} \mathbf{E}(X_i)^2 - (\mathbf{E}X)^2 \\
&= \sum_{i\neq j} \mathbf{E}(X_i X_j) + \sum_{i=i}^{n} \mathbf{E}X_i - (\mathbf{E}X)^2 \\
&= 2\binom{n}{2}(1-p)^{n-2}(1-p)^{n-2}(1-p) + \mathbf{E}X - (\mathbf{E}X)^2 \\
&= n(n-1)(1-p)^{2n-3} + \mathbf{E}X - (\mathbf{E}X)^2
\end{aligned}
$$

since

$$\mathbf{P}(X=0) \le \mathbf{P}(|X-\mathbf{E}X| \ge \mathbf{E}X) \le \frac{\mathrm{Var}X}{(\mathbf{E}X)^2}$$

thus

$$\begin{aligned}
\mathbf{P}(X=0) &\le \frac{n(n-1)(1-p)^{2n-3}+\mathbf{E}X-(\mathbf{E}X)^2}{(\mathbf{E}X)^2} \\
&= \frac{n(n-1)(1-p)^{2n-3}+n(1-p)^{n-1}-n^2(1-p)^{2n-2}}{n^2(1-p)^{2n-2}} \\
&\le \frac{n^2(1-p)^{2n-3}p+n(1-p)^{n-1}}{n^2(1-p)^{2n-2}} \\
&= \frac{p}{1-p}+\frac{1}{n(1-p)^{n-1}}
\end{aligned}$$

if $p = c\frac{\ln n}{n}$, $c < 1$, $\epsilon > 0$ and $n \to \infty$, $\mathbf{P}(X=0) \to 0$
the graph has isolated vertices with probability at least $1-\epsilon$