CS787: Advanced Algorithms	Scribe: Xinyi Li
Lecture 8: Concentration	**Date:** 02/19/19

Outlines for the Following Lecture

- Concentration Bounds
- Routing to minimize congestion

8.1 Introduction of Some Inequalities

8.1.1 Markov's Inequality

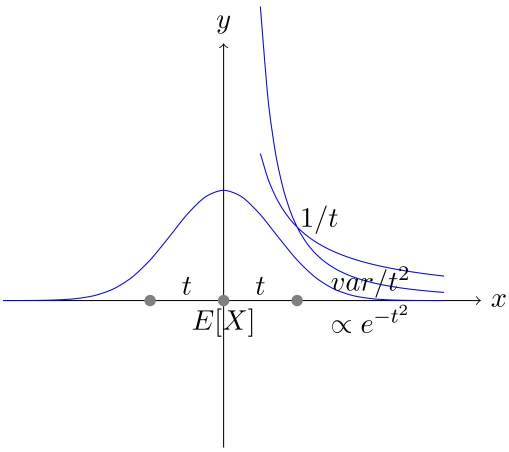
Theorem 8.1.1 *For any non-negative random variable x, and any $t \geqslant 0$:*

$$Pr[x \geqslant t] \leqslant \frac{E[x]}{t} \tag{8.1.1}$$

8.1.2 Chebyshev's Inequality

Theorem 8.1.2 *For any non-negative random variable x, and any $t \geqslant 0$:*

$$Pr[|x - E[x]| \geqslant t] \leqslant \frac{\sigma^2(x)}{t^2} \tag{8.1.2}$$



Example: What is $Pr[\#heads \geqslant \frac{3n}{4}]$ if flipping a fair coin for n times?
Let x_i be the random variable for ith flipping result being head. $X = \sum_i x_i$ is the variable to represent the number of heads after n times flipping. $E[X] = \frac{n}{2}$ and $\sigma(X) = \frac{n}{4}$. (For independent variables x_is, $\sigma^2(\sum_i x_i) = \sum_i \sigma^2(x_i)$)
Markov's Inequality gives $Pr[\#heads \geqslant \frac{3n}{4}] \leqslant \frac{n/2}{3n/4} = \frac{2}{3}$

Chebyshev's Inequality gives $Pr[\#heads \geqslant \frac{3n}{4}] \leqslant \frac{n/4}{(n/4)^2} = \frac{4}{n}$
Chebyshev's Inequality gives a much tight probability estimation compared to Markov's Inequality. And the probability decreases with increase of flipping times n. ■

8.1.3 Chernoff-Hoeffding Inequality

Theorem 8.1.3 *Let $x_1, x_2 \cdots, x_n$ be independent and bounded variables i.e. $x_i \in [0,1]$, $\forall i \in n$, let $X = \sum_i x_i$ and $\mu = E[X]$. Then for any $\delta > 0$*

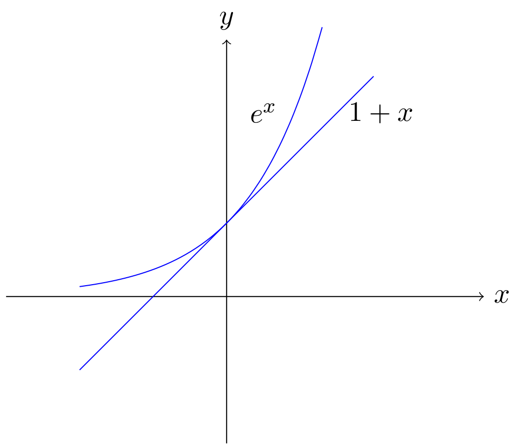
$$Pr[x \geqslant (1+\delta)\mu] \leqslant \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu} \tag{8.1.3}$$

$$Pr[x \geqslant (1-\delta)\mu] \leqslant \left[\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right]^{\mu} \tag{8.1.4}$$

We are proving the bound for the special case of $x_i \in \{0,1\}$ although it holds also for the more general case stated.
Proof: Assume $x_i \in \{0,1\}$, $Pr(x_i = 1) = P_i$ and $\mu = \sum_i P_i$. For value t,

$$\begin{aligned} E[e^{x_i t}] &= P_i e^t + (1 - P_i) \\ &= 1 + P_i(e^t - 1) \leqslant e^{P_i(e^t-1)} \end{aligned} \tag{8.1.5}$$

The above formula is based on inequality $(1+x) \leqslant e^x, \forall x$, which can be shown in the following graph



Let $f(x)$ be any non-negative increasing function. Then Markov's Inequality gives us

$$Pr[x \geqslant \lambda] = Pr[f(x) \geqslant f(\lambda)] \leqslant \frac{E[f(x)]}{f(\lambda)} \tag{8.1.6}$$

Now choose $f(x) = e^{xt}$ for some $t > 0$,

$$\begin{aligned} Pr[x \leqslant (1+\delta)\mu] &\leqslant \frac{E(e^{xt})}{exp((1+\delta)\mu t)} = \frac{E(e^{\sum_i x_i t})}{e^{(1+\delta)\mu t}} = \frac{\prod_i E(e^{x_i t})}{e^{(1+\delta)\mu t}} \\ &= \frac{\prod_i \{1 + P_i(e^t - 1)\}}{e^{(1+\delta)\mu t}} \leqslant \frac{\prod_i e^{P_i(e^t-1)}}{e^{(1+\delta)\mu t}} = \frac{e^{\mu(e^t-1)}}{e^{\mu(1+\delta)t}} \\ &= \left[\frac{e^{e^t-1}}{e^{(1+\delta)t}}\right]^\mu \end{aligned} \tag{8.1.7}$$

$\left[\frac{e^{e^t-1}}{e^{(1+\delta)t}}\right]^\mu$ gets the minimum value $\left[\frac{e^\delta}{(1+\delta)^{1+\delta}}\right]^\mu$ when $t = \ln(1+\delta)$. Thus inequality $Pr[x \geqslant (1+\delta)\mu] \leqslant \left[\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right]^\mu$ is proved. To prove $Pr[x \geqslant (1-\delta)\mu] \leqslant \left[\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right]^\mu$, we can choose a decreasing function $f(x) = e^{-xt}$ and apply the same proof. ■

Chernoff-Hoeffding Inequality implies other inequalities:
if $\delta > 0$,

$$Pr[x \geqslant (1+\delta)\mu] \leqslant exp\{-\frac{\delta^2}{2+\delta}\mu\} \tag{8.1.8}$$

if $0 \leqslant \delta \leqslant 1$,

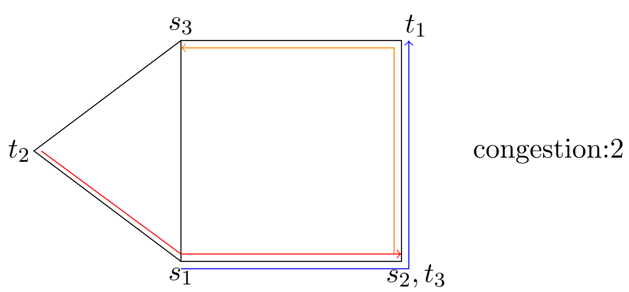
$$Pr[x \geqslant (1+\delta)\mu] \leqslant exp\{-\frac{\delta^2}{3}\mu\} \tag{8.1.9}$$

$$Pr[x \leqslant (1-\delta)\mu] \leqslant exp\{-\frac{\delta^2}{2}\mu\} \tag{8.1.10}$$

For the coin example, $\mu = \frac{n}{2}$ and $\delta = \frac{1}{2}$, $Pr[\#heads \geqslant \frac{3n}{4}] \leqslant e^{-n/24}$.
For what value of t, is $Pr[\#heads \geqslant t] \leqslant \frac{1}{n}$?
When $e^{-\delta^2 n/6} = \frac{1}{n}$, $\delta = \sqrt{\frac{6\log n}{n}}$. The probability of # heads larger than $\frac{n}{2} + \frac{n}{2}\sqrt{\frac{6\log n}{n}} \approx \frac{n}{2} + \sqrt{n\log n}$ is small.

8.2 Routing to Minimize Congestion

Given Graph $G = (V, E)$ and k source-sink pairs (s_i, t_i), find paths P_i from s_i to t_i. Define the congestion of edge e as # of paths it belongs to($|\{i : P_i \ni e\}|$). Our goal is to minimize the maximum congestion among all edges($\min \max_{e\in E} congestion(e)$).



8.2.1 ILP Formulation

Let P_i be the set of all paths in G from s_i to t_i and $x_{i,P}$ be the integer random variable for choosing path $P \in P_i$. t represents congestion. Then the problem can be formulate as the following integer linear programming:

$$\begin{aligned} minimize\ t, s.t. \sum_i \sum_{P \in P_i, P \ni e} x_{i,P} \leqslant t, \forall e \\ x_{i,p} \in \{0,1\}, \forall i, P \in P_i \\ \sum_{P \in P_{in}} x_{i,P} = 1, \forall i \end{aligned} \tag{8.2.11}$$

We can relax this problem to LP problem by letting $x_{i,p} \in [0,1]$. But the number of potential paths between any $s_i - t_i$ pair is exponential. Instead, we employ a different LP formula using edge variables. Let $y_{i,e}$ be the flow of commodity of i on edge e:

$$\begin{aligned} minimize\ t, s.t. \sum_i y_{i,e} \leqslant t, \forall e \\ y_{i,e} \in [0,1], \forall i, e \\ \sum_{e \in \delta^-(v)} y_{i,e} = \sum_{e \in \delta^+(v)} y_{i,e}, \forall i, v \in V \setminus \{s_i, t_i\} \\ \sum_{e \in \delta^-(s_i)} y_{i,e} = \sum_{e \in \delta^+(t_i)} y_{i,e} = 1, \forall i \end{aligned} \tag{8.2.12}$$

8.2.2 Rounding Algorithms

To perform randomized rounding, the fractional edge-based flows obtained by solving the above LP are converted into path-based flows for each $s_i - t_i$ pair using a standard flow decomposition.

Algorithm 1 Flow Decomposition

Step 1: Start with some (s_i, t_i) flow f
Step 2: Find some flow carrying path $p \in P_i$
Step 3: Assign as much flow to p as possible
Step 4: Remove p from$f$
Step 5: Repeat step4 until all paths are deleted

For each i, this gives a collection of $s_i - t_i$ paths, with each path assigned a fractional weight. The sum of weights over all paths for each i is 1. Randomized rounding is performed by selecting a single path for each i based on the weight over all $s_i - t_i$ paths.