Randomized Algorithms

The Chernoff bound

Speaker: Chuang-Chieh Lin
Advisor: Professor Maw-Shang Chang
National Chung Cheng University
2006/10/25

Outline

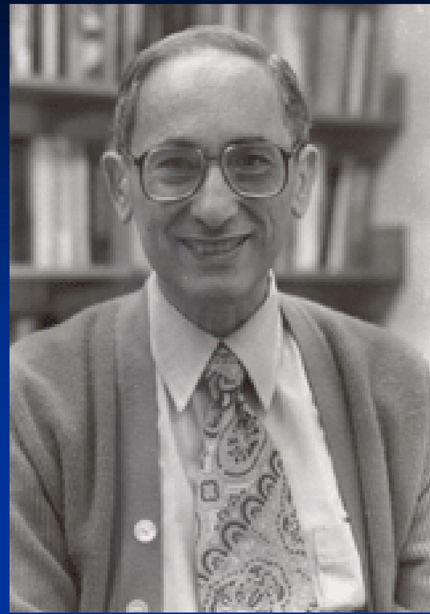
- Introduction
- The Chernoff bound
- Markov's Inequality
- The Moment Generating Functions
- The Chernoff Bound for a Sum of Poisson Trials
- The Chernoff Bound for Special cases
- Set Balancing Problem
- Error-reduction for **BPP**
- References

Introduction

- Goal:
  - *The Chernoff bound* can be used in the analysis on the *tail of the distribution* of the ***sum of independent random variables***, with some extensions to the case of dependent or correlated random variables.

- *Markov's Inequality* and *Moment generating functions* which we shall introduce will be greatly needed.

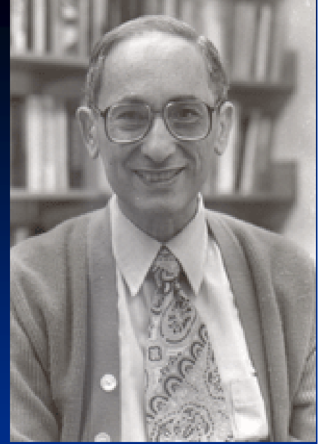
Math tool



Professor Herman Chernoff's bound,

Annal of Mathematical Statistics 1952

Chernoff bounds



In it′s most general form, the *Chernoff bound for a random variable* X is obtained as follows: for any $t > 0$,

$$\mathbf{Pr}[X \geq a] \leq \frac{\mathbf{E}[e^{tX}]}{e^{ta}}$$

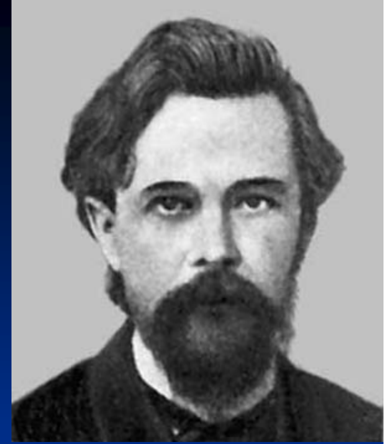
A moment generating function

or equivalently,

$$\ln \mathbf{Pr}[X \geq a] \leq -ta + \ln \mathbf{E}[e^{tX}].$$

The value of t that minimizes $\dfrac{\mathbf{E}[e^{tX}]}{e^{ta}}$ gives the best possible bounds.

Markov's Inequality



For any random variable $X \geq 0$ and any $a > 0$,

$$\mathbf{Pr}[X \geq a] \leq \frac{\mathbf{E}[X]}{a}.$$

We can use Markov's Inequality to derive the famous *Chebyshev's Inequality*:

$$\mathbf{Pr}[|X - \mathbf{E}[X]| \geq a] = \mathbf{Pr}[(X - \mathbf{E}[X])^2 \geq a^2] \leq \frac{\mathbf{Var}[X]}{a^2}.$$

Proof of the Chernoff bound

It follows directly from Markov's inequality:

$$\mathbf{Pr}[X \geq a] = \mathbf{Pr}[e^{tX} \geq e^{ta}]$$

$$\leq \frac{\mathbf{E}[e^{tX}]}{e^{ta}}$$

So, how to calculate this term?

Moment Generating Functions

$$M_X(t) = \mathbf{E}[e^{tX}].$$

This function gets its name because we can generate the ith moment by differentiating $M_X(t)$ i times and then evaluating the result for $t = 0$:

$$\left.\frac{d^i}{dt^i} M_X(t)\right|_{t=0} = \mathbf{E}[X^i].$$

The ith moment of r.v. X

Remark: $\mathbf{E}[X^i] = \sum_{x \in X} x^i \cdot \mathbf{Pr}[X = x]$

Moment Generating Functions (cont'd)

We can easily see why the moment generating function works as follows:

$$
\begin{aligned}
\frac{d^i}{dt^i} M_X(t)\Big|_{t=0} &= \frac{d^i}{dt^i}\mathbf{E}[e^{tX}]\Big|_{t=0} \\
&= \frac{d^i}{dt^i}\sum_s e^{ts}\mathbf{Pr}[X=s]\Big|_{t=0} \\
&= \sum_s \frac{d^i}{dt^i} e^{ts}\mathbf{Pr}[X=s]\Big|_{t=0} \\
&= \sum_s s^i e^{ts}\mathbf{Pr}[X=s]\Big|_{t=0} \\
&= \sum_s s^i \mathbf{Pr}[X=s] \\
&= \mathbf{E}[X^i].
\end{aligned}
$$

Moment Generating Functions (cont'd)

- The concept of the moment generating function (mgf) is connected with a distribution rather than with a random variable.

- Two different random variables with the same distribution will have the same mgf.

Moment Generating Functions (cont'd)

- **Fact:** If $M_X(t) = M_Y(t)$ for all $t \in (-c, c)$ for some $c > 0$, then X and Y have the same distribution.

- If X and Y are two independent random variables, then

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

- Let $X_1, \ldots, X_k$ be independent random variables with mgf's $M_1(t), \ldots, M_k(t)$. Then the mgf of the random variable $Y = \sum_{i=1}^{k} X_i$ is given by

$$M_Y(t) = \prod_{i=1}^{k} M_i(t).$$

Moment Generating Functions (cont'd)

- If X and Y are two independent random variables, then

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

Proof:

$$\begin{aligned} M_{X+Y}(t) &= \mathbf{E}[e^{t(X+Y)}] \\ &= \mathbf{E}[e^{tX}e^{tY}] \\ &= \mathbf{E}[e^{tX}]\mathbf{E}[e^{tY}] \\ &= M_X(t)M_Y(t). \end{aligned}$$

Here we have used that X and Y are independent – and hence e^{tX} and e^{tY} are independent – to conclude that $\mathbf{E}[e^{tX}e^{tY}] = \mathbf{E}[e^{tX}]\mathbf{E}[e^{tY}]$.

Chernoff bound for the sum of Poisson trials

- Poisson trials:
  - The distribution of a sum of independent 0-1 random variables, which may not be identical.
- Bernoulli trials:
  - The same as above except that all the random variables are identical.

Chernoff bound for the sum of Poisson trials (cont'd)

★ $X_i : i = 1, \ldots, n$, mutually independent 0-1 random variables with $\mathbf{Pr}[X_i = 1] = p_i$ and $\mathbf{Pr}[X_i = 0] = 1 - p_i$.

Let $X = X_1 + \ldots + X_n$ and $\mathbf{E}[X] = \mu = p_1 + \ldots + p_n$.

$M_{X_i}(t) = \mathbf{E}[e^{tX_i}] = p_i e^{t \cdot 1} + (1 - p_i)e^{t \cdot 0} = p_i e^t + (1 - p_i)$

$= 1 + p_i(e^t - 1) \leq e^{p_i(e^t - 1)}$. (Since $1 + y \leq e^y$.)

★ $M_X(t) = \mathbf{E}[e^{tX}] = M_{X_1}(t) M_{X_2}(t) \ldots M_{X_n}(t) \leq e^{(p_1 + p_2 + \ldots + p_n)(e^t - 1)}$

$= e^{(e^t - 1)\mu}$,

since $\mu = p_1 + p_2 + \ldots + p_n$.

We will use this result later.

Chernoff bound for the sum of Poisson trials (cont'd)

Poisson trials

Theorem 1: Let $X = X_1 + \cdots + X_n$, where $X_1, \ldots, X_n$ are n independent trials such that $\mathbf{Pr}[X_i = 1] = p_i$ holds for each $i = 1, 2, \ldots, n$. Then,

(1) for any $d > 0$, $\mathbf{Pr}[X \geq (1+d)\mu] \leq \left(\frac{e^d}{(1+d)^{1+d}}\right)^{\mu}$;

(2) for $d \in (0, 1]$, $\mathbf{Pr}[X \geq (1+d)\mu] \leq e^{-\mu d^2/3}$;

(3) for $R \geq 6\mu$, $\mathbf{Pr}[X \geq R] \leq 2^{-R}$.

Proof of Theorem 1:

For any random variable $X \geq 0$ and any $a > 0$, $\mathbf{Pr}[X \geq a] \leq \frac{\mathbf{E}[X]}{a}$.

By Markov inequality, for any $t > 0$ we have

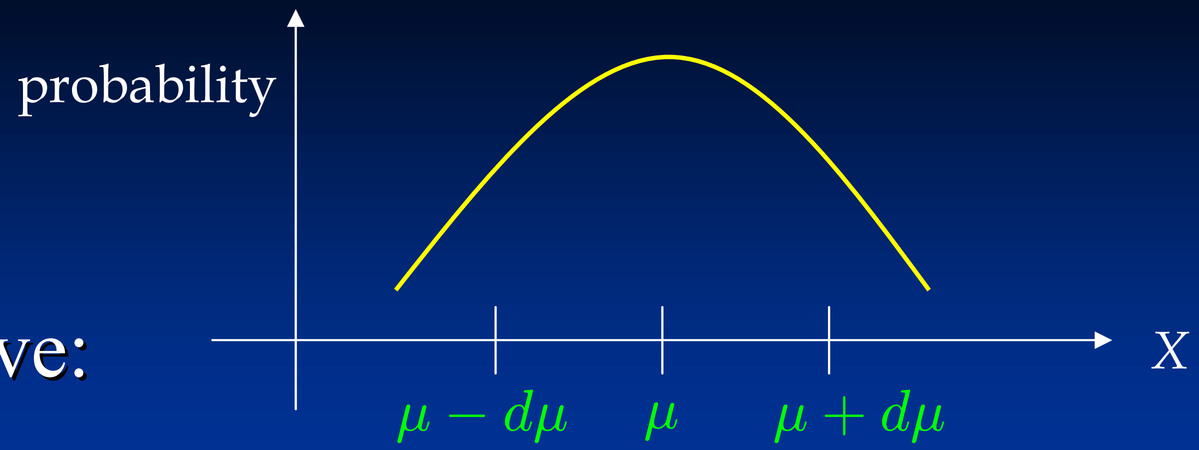
$\mathbf{Pr}[X \geq (1+d)\mu] = \mathbf{Pr}[e^{tX} \geq e^{t(1+d)\mu}] \leq \mathbf{E}[e^{tX}]/e^{t(1+d)\mu} \leq e^{(e^t-1)\mu}/e^{t(1+d)\mu}$. For any $d > 0$, set $t = \ln(1+d) > 0$ we have (1).

To prove (2), we need to show for $0 < d \leq 1$, $e^d/(1+d)^{(1+d)} \leq e^{-d^2/3}$.
Taking the logarithm of both sides, we have $d-(1+d)\ln(1+d) + d^2/3 \leq 0$, which can be proved with calculus.

To prove (3), let $R = (1+d)\mu$. Then, for $R \geq 6\mu$, $d = R/\mu - 1 \geq 5$. Hence, using (1), $\mathbf{Pr}[X \geq (1+d)\mu] \leq \left(\frac{e^d}{(1+d)^{(1+d)}}\right)^{\mu} \leq \left(\frac{e}{1+d}\right)^{(1+d)\mu} \leq (e/6)^R \leq 2^{-R}$.

from (1)

- Similarly, we have:



Theorem: Let $X = \sum_{i=1}^{n} X_i$, where $X_1, \ldots, X_n$ are n independent Poisson trials such that $\mathbf{Pr}[X_i = 1] = p_i$. Let $\mu = \mathbf{E}[X]$. Then, for $0 < d < 1$:

(1) $\mathbf{Pr}[X \le (1-d)\mu] \le \left(\frac{e^{-d}}{(1-d)^{(1-d)}}\right)^{\mu}$;

(2) $\mathbf{Pr}[X \le (1-d)\mu] \le e^{-\mu d^2/2}$.

Corollary: For $0 < d < 1$, $\mathbf{Pr}[|X - \mu| \ge d\mu] \le 2e^{-\mu d^2/3}$.

- Example: Let *X* be the number of heads of *n* independent fair coin flips. Applying the above Corollary, we have:

$$\Pr[|X - n/2| \geq \sqrt{6n \ln n}/2] \leq 2\exp(-\frac{1}{3}\frac{n}{2}\frac{6\ln n}{n}) = 2/n.$$

$$\Pr[|X - n/2| \geq n/4] \leq 2\exp(-\frac{1}{3}\frac{n}{2}\frac{1}{4}) = 2e^{-n/24}.$$ → **Better!!**

By Chebyshev's inequality, i.e. $\mathbf{Pr}[|X - \mathbf{E}[X]| \geq a] \leq \frac{\mathbf{Var}[X]}{a^2}$, we have $\Pr[|X - n/2| \geq n/4] \leq 4/n$.

Better bounds for special cases

Theorem Let $X = X_1 + \cdots + X_n$, where $X_1, \ldots, X_n$ are n independent random variables with $\mathbf{Pr}[X_i = 1] = \mathbf{Pr}[X_i = -1] = 1/2$. For any $a > 0$, $\mathbf{Pr}[X \geq a] \leq e^{-a^2/2n}$.

Proof: For any $t > 0$, $\mathbf{E}[e^{tX_i}] = e^{t\cdot 1}/2 + e^{t\cdot(-1)}/2$.

Since $e^t = 1 + t + t^2/2! + \cdots + t^i/i! + \cdots$ and $e^{-t} = 1 - t + t^2/2! + \cdots + (-1)^i t^i/i! + \cdots$, using Taylor series, we have

$\mathbf{E}[e^{tX_i}] = \sum_{i\geq 0} t^{2i}/(2i)! \leq \sum_{i\geq 0}(t^2/2)^i/i! = e^{t^2/2}$.

$\mathbf{E}[e^{tX}] = \prod_{i=1}^{n} \mathbf{E}[e^{tX_i}] \leq e^{t^2 n/2}$ and $\mathbf{Pr}[X \geq a] = \mathbf{Pr}[e^{tX} \geq e^{ta}] \leq \mathbf{E}[e^{tX}]/e^{ta} \leq e^{t^2 n/2}/e^{ta}$. Setting $t = a/n$, we have $\mathbf{Pr}[X \geq a] \leq e^{-a^2/2n}$. By symmetry, we have $\mathbf{Pr}[X \leq -a] \leq e^{-a^2/2n}$.

Better bounds for special cases (cont'd)

Corollary Let $X = X_1 + \cdots + X_n$, where $X_1, \ldots, X_n$ are n independent random variables with $\mathbf{Pr}[X_i = 1] = \mathbf{Pr}[X_i = -1] = 1/2$. For any $a > 0$, $\mathbf{Pr}[|X| \geq a] \leq 2e^{-a^2/2n}$.

Let $Y_i = (X_i + 1)/2$, we have the following corollary.

Better bounds for special cases (cont'd)

Corollary Let $Y = Y_1 + \cdots + Y_n$, where $Y_1, \ldots, Y_n$ are n independent random variables with $\mathbf{Pr}[Y_i = 1] = \mathbf{Pr}[Y_i = 0] = 1/2$. Let $\mu = \mathbf{E}[Y] = n/2$.

(1) For any $a > 0$, $\mathbf{Pr}[Y \geq \mu + a] \leq e^{-2a^2/n}$.

(2) For any $d > 0$, $\mathbf{Pr}[Y \geq (1+d)\mu] \leq e^{-d^2\mu}$.

(3) For any $\mu > a > 0$, $\mathbf{Pr}[Y \leq \mu - a] \leq e^{-2a^2/n}$.

(4) For any $1 > d > 0$, $\mathbf{Pr}[Y \leq (1-d)\mu] \leq e^{-d^2\mu}$.

Note: *The details can be left for exercises. (See* [MU05], *pp.* 70-71.)

An application: Set Balancing

- Given an $n \times m$ matrix $\mathbf{A}$ with entries in $\{0,1\}$, let

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

- Suppose that we are looking for a vector $\mathbf{v}$ with entries in $\{-1, 1\}$ that *minimizes*

$$\| \mathbf{Av} \|_\infty = \max_{i=1,\ldots,n} |c_i|.$$

Set Balancing (cont'd)

- The problem arises in designing statistical experiments.
- Each column of matrix **A** represents a subject in the experiment and each row represents a feature.
- The vector **v** partitions the subjects into two disjoint groups, so that each feature is roughly as balanced as possible between the two groups.

Set Balancing (cont'd)

For example,

A:

	斑馬	老虎	鯨魚	企鵝
肉食性	0	1	0	0
陸生	1	1	0	0
哺乳類	1	1	1	0
產卵	0	0	0	1

v:

1
1
−1
−1

Av:

1
2
1
−1

We obtain that $\| \mathbf{Av} \|_\infty = 2$.

Set Balancing (cont'd)

For example,

A:

	斑馬	老虎	鯨魚	企鵝
肉食性	0	1	0	0
陸生	1	1	0	0
哺乳類	1	1	1	0
產卵	0	0	0	1

v:

–1
1
1
–1

Av:

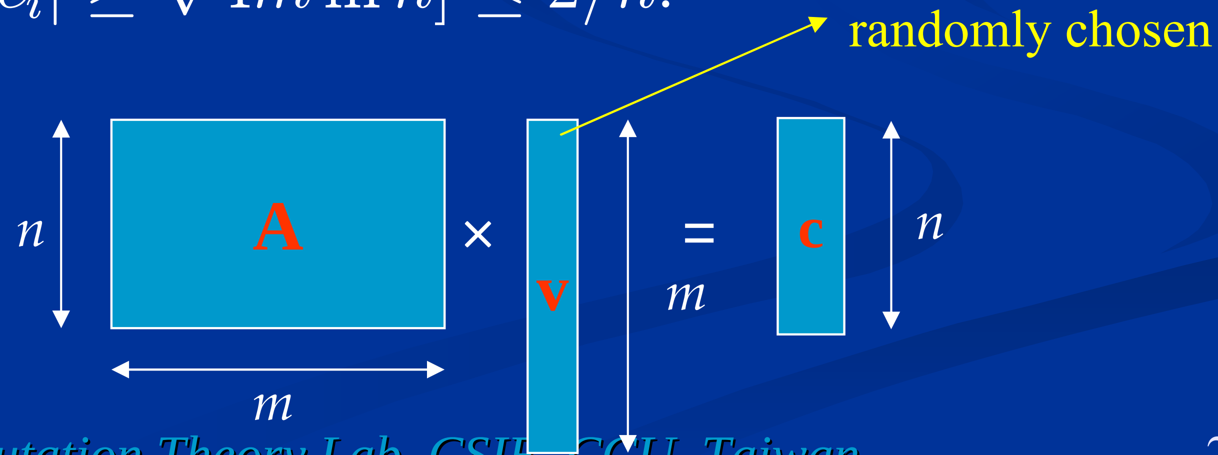
1
0
1
–1

We obtain that $\| \mathbf{Av} \|_\infty = 1$.

Set Balancing (cont'd)

Set balancing: Given an $n \times m$ matrix $\mathbf{A}$ with entries 0 or 1, let $\mathbf{v}$ be an m-dimensional vector with entries in $\{1, -1\}$ and $\mathbf{c}$ be an n-dimensional vector such that $\mathbf{A}\mathbf{v} = \mathbf{c}$.

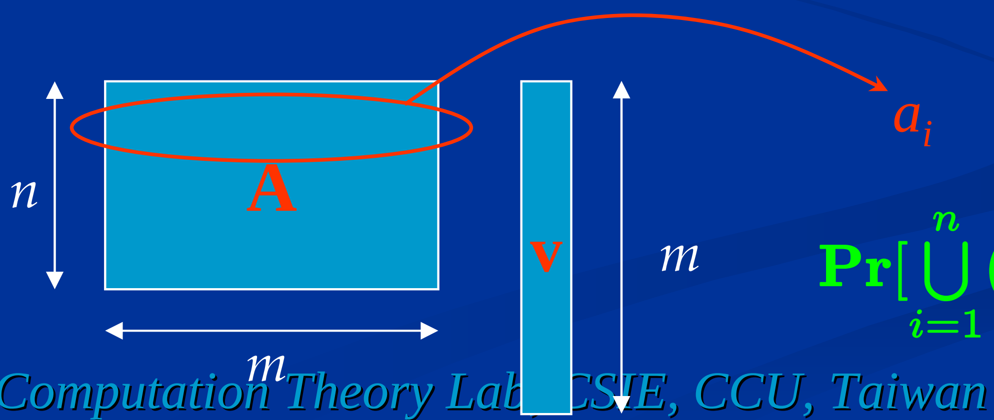
Theorem For a random vector $\mathbf{v}$ with entries chosen randomly and with equal probability from the set $\{1, -1\}$, $\mathbf{Pr}[\max_i |c_i| \geq \sqrt{4m \ln n}] \leq 2/n$.



Proof of Set Balancing:

<u>**Proof:**</u> Consider the i-th row of $\mathbf{A}$: $a_i = (a_{i,1}, \cdots, a_{i,m})$. Suppose there are k 1s in a_i. If $k < \sqrt{4m \ln n}$, then clearly $|a_i \mathbf{v}| \le \sqrt{4m \ln n}$. Suppose $k \ge \sqrt{4m \ln n}$, then there are k non-zero terms in $Z_i = \sum_{j=1}^{m} a_{i,j} v_j$, which are independent random variables, each with probability $1/2$ of being either $+1$ or -1.

By the Chernoff bound and the fact $m \ge k$, we have $\mathbf{Pr}[|Z_i| \ge \sqrt{4m \ln n}] \le 2e^{-4m \ln n / 2k} \le 2/n^2$. By the union bound we have the bound for every row is at most $2/n$.



$$\mathbf{Pr}[\bigcup_{i=1}^{n} (|Z_i| > \sqrt{4m \ln n})]$$

Another application: Error-reduction in BPP

- The class BPP (for Bounded-error Probabilistic Polynomial time) consists of all languages L that have a randomized algorithm A running in worst-case polynomial time that for any input $x \in \Sigma^*$,
  - $x \in L \Rightarrow \mathbf{Pr}[A(x) \text{ accepts}] \geq 3/4$.
  - $x \notin L \Rightarrow \mathbf{Pr}[A(x) \text{ rejects}] \geq 3/4$.

That is, the error probability is at most ¼.

Error-reduction in BPP (cont'd)

- Consider the following variant definition:
- The class **BPP** (for Bounded-error Probabilistic Polynomial time) consists of all languages L that have a randomized algorithm A running in worst-case polynomial time that for any input $x \in \Sigma^*$ with $|x| = n$ and some positive integer $k \geq 2$,
  - $x \in L \Rightarrow \mathbf{Pr}[A(x) \text{ accepts}] \geq ½ + n^{-k}$.
  - $x \notin L \Rightarrow \mathbf{Pr}[A(x) \text{ rejects}] \geq ½ + n^{-k}$.

Error-reduction in BPP (cont'd)

- The previous two definitions of **BPP** are equivalent.
- We will show that the latter one can be transferred to the former one by Chernoff bounds as follows.
- Let M_A be an algorithm simulating algorithm A for "***t***" times and output the majority answer.
  - That is, if there are more than $t/2$ "accepts", M_A will output "Accept".
  - Otherwise, M_A will output "Reject".

Error-reduction in BPP (cont'd)

- Let X_i, for $1 \le i \le t$, be a random variable such that $X_i = 1$ if the ith execution of M_A (running algorithm A) produces a *correct* answer and $X_i = 0$ otherwise.
  - That is, accepts if $x \in L$ and rejects if $x \notin L$.

- Let $X = \sum_{i=1}^{t} X_i$, we have $\mu_X \ge (\frac{1}{2} + \frac{1}{n^k})t = t \cdot \frac{n^k+2}{2n^k}$.

  So $\frac{t}{2} \le \frac{n^k}{n^k+2} \cdot \mu_X$.

Error-reduction in BPP (cont'd)

- Recall one of the previous results of the Chernoff bound:

Theorem: Let $X = \sum_{i=1}^{n} X_i$, where $X_1, \ldots, X_n$ are n independent Poisson trials such that $\mathbf{Pr}[X_i = 1] = p_i$. Let $\mu = \mathbf{E}[X]$. Then, for $0 < d < 1$:

(1) $\mathbf{Pr}[X \leq (1-d)\mu] \leq \left(\frac{e^{-d}}{(1-d)^{(1-d)}}\right)^{\mu}$;

(2) $\mathbf{Pr}[X \leq (1-d)\mu] \leq e^{-\mu d^2/2}$.

Error-reduction in BPP (cont'd)

- We have the error probability

$$\begin{aligned} \Pr[X < t/2] &\leq \Pr[X < \frac{n^k}{n^k+2} \cdot \mu_X] \\ &\leq \Pr[X \leq \left(1 - \frac{2}{n^k+2}\right)\mu_X] \\ &\leq e^{-\mu_X(\frac{2}{n^k+2})^2/2} \\ &= e^{-\mu_X \frac{2}{(n^k+2)^2}} \\ &\leq e^{-\frac{t}{n^k(n^k+2)}}. \end{aligned}$$

- Let $e^{-\frac{t}{n^k(n^k+2)}} \leq 1/4$, we can derive that the value of t as follows.

Error-reduction in BPP (cont'd)

- By taking logarithm on both sides, we have

$$-\frac{t}{n^k(n^k+2)} \leq \ln\frac{1}{4}$$

So we can take t to be $\ln 4 \cdot n^k(n^k+2)$, then we have

$$\begin{aligned}
\Pr[X < t/2] &\leq e^{-\frac{t}{n^k(n^k+2)}} \\
&= e^{-\frac{\ln 4 \cdot n^k(n^k+2)}{n^k(n^k+2)}} \\
&= e^{-\ln 4} \\
&= 1/4.
\end{aligned}$$

Error-reduction in BPP (cont'd)

- Since $t = \ln 4 \cdot (n^{2k} + 2n^{k})$ is still polynomial, the running time of M_A will be still polynomial. Hence the latter definition for **BPP** is equivalent to the former one!

References

- [MR95] Rajeev Motwani and Prabhakar Raghavan, *Randomized algorithms*, Cambridge University Press, 1995.
- [MU05] Michael Mitzenmacher and Eli Upfal, *Probability and Computing - Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, 2005.
- 蔡錫鈞教授上課投影片
- Professor Valentine Kabanets's lectures

Thank you.