Probabilistic Method and Random Graphs

Lecture 3. Chernoff bounds: behind and beyond

Xingwu Liu

Institute of Computing Technology
Chinese Academy of Sciences, Beijing, China

Preface

Questions, comments, or suggestions?

A brief review

Moments

Expectation, k-moment, variance

Inequalities

Universal: Union bound
1-moment: Markov's inequality
2-moment: Chebychev's inequality

Chernoff bounds: independent sum

Let $X = \sum_{i=1}^{n} X_i$, where $X_i' s$ are **independent** Poisson trials. Let $\mu = \mathbb{E}[X]$. Then

1. For $\delta > 0$, $\Pr(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu} \leq e^{-\frac{\delta^2}{2+\delta}\mu}$.
2. For $1 > \delta > 0$, $\Pr(X \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu} \leq e^{-\frac{\delta^2}{2}\mu}$.

General bounds for independent sums

Each $X_i \in [0,1]$ but is not necessarily a Poisson trial

Basic Chernoff bounds remain valid (by $e^{\lambda x} \leq x e^{1\lambda} + (1-x)e^{0\lambda}$).

Each $X_i \in [0, s]$

Basic Chernoff bounds remain valid, except that the exponent μ is divided by s.

The domains (a_i, b_i) of $X_i's$ differ

Hoeffding's Inequality: $\Pr(|X - \mathbb{E}[X]| \geq t) \leq 2e^{-\frac{2t^2}{\sum_i (b_i - a_i)^2}}$.
Proposed in 1963.

Remarks of Hoeffding's Inequality

1. It considers the absolute, rather than relative, deviation. Particularly useful if $\mu = 0$.
2. When each $X_i \in [0, s]$, it is tighter than the simplified basic Chernoff bounds if δ is big, and looser otherwise.

Hoeffding's Inequality

Let $X = \sum_{i=1}^{n} X_i$, where $X_i \in [a_i, b_i]$ are independent r.v. Then $\Pr(|X - \mathbb{E}[X]| \geq t) \leq 2e^{-\frac{2t^2}{\sum_i (b_i - a_i)^2}}$ for any $t > 0$

Idea of the proof

1. Given r.v. $Z \in [a, b]$ with $\mathbb{E}[Z] = 0$, $\mathbb{E}[e^{\lambda Z}] \leq e^{\frac{\lambda^2 (b-a)^2}{8}}$.
2. $\Pr(X - \mathbb{E}[X] \geq t) \leq \frac{\prod_i \mathbb{E}[e^{\lambda(X_i - \mathbb{E}[X_i])}]}{e^{\lambda t}} \leq e^{\lambda^2 \sum_i \frac{(b_i - a_i)^2}{8} - \lambda t}$

Proof of Fact 1

1. $e^{\lambda z} \leq \frac{z-a}{b-a} e^{\lambda b} + \frac{b-z}{b-a} e^{\lambda a}$, for $z \in [a, b]$.
2. $\mathbb{E}[e^{\lambda Z}] \leq (1 - \theta + \theta e^{u}) e^{-\theta u}$ where $\theta = \frac{-a}{b-a}$ and $u = \lambda(b - a)$.
3. Define $\phi(x) \triangleq -\theta x + \ln(1 - \theta + \theta e^{x})$. Then $\mathbb{E}[e^{\lambda Z}] \leq e^{\phi(u)}$.
4. Use calculus to show that $\phi(u) \leq \frac{u^2}{8}$

Example: Hoeffding's Inequality + Union bound

Set balancing

Given a matrix $A \in \{0,1\}^{n \times m}$, find $b \in \{-1,1\}^m$ s.t. $\| Ab \|_\infty$ is minimized.

Motivation

$$\begin{matrix} \text{feature 1:} \\ \text{feature 2:} \\ \vdots \\ \text{feature n:} \end{matrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}, \text{ each column is an object.}$$

Want to partition the objects so that every feature is balanced.

Algorithm

Uniformly randomly sample b.

Performance analysis

Performance

$\Pr(\| Ab \|_\infty \geq \sqrt{4m \ln n}) \leq \frac{2}{n}$

Proof

For any $1 \leq i \leq n$, $Z_i = \sum_j a_{ij} b_j$ is the ith entry of Ab. By union bound, it suffices to prove $\Pr(|Z_i| \geq \sqrt{4m \ln n}) \leq \frac{2}{n^2}$ for each i.

Fix i. W.l.o.g, assume $a_{ij} = 1$ iff $1 \leq j \leq k$ for some $k \leq m$. Then $Z_i = b_1 + ... + b_k$.

The b_j's are independent over $\{-1, 1\}$ with $\mathbb{E}[b_j] = 0$.

By Hoeffding's Inequality, $\Pr(|Z_i| \geq \sqrt{4m \ln n}) \leq 2e^{-\frac{8m \ln n}{4k}} \leq \frac{2}{n^2}$

Reflection on moments and Chernoff bounds

Moments

Do moments uniquely determine the distribution?

Chernoff Bounds

Why is it so good?
Can it be improved by non-exponential functions?
Anything to do with moments?

The story begins with generating functions.

Generating functions

Informal definition

A power series whose coefficients encode information about a sequence of numbers.

Example: Probability generating function

Given a discrete random variable X whose values are non-negative integers, $G_X(t) \triangleq \sum_{n\geq 0} t^n \Pr(X = n) = \mathbb{E}[t^X]$.
Example: a Bernoulli random variable.

Properties

Convergence: It converges if $|t| < 1$.
Uniqueness: $G_X(\cdot) \equiv G_Y(\cdot)$ implies the same distribution.

Application

Toy: Use uniqueness to show that the summation of independent identical binomial distribution is binomial.
Deriving Moments: $G_X^{(k)}(1) = \mathbb{E}[X(X-1)\cdots(X-k+1)]$.

Moment generating functions

Shortcoming of probability generating functions

Only valid for non-nagetive integer random variables.

Moment generating functions

$M_X(t) \triangleq \sum_x e^{tx} \Pr(X = x) = \mathbb{E}[e^{tX}]$.
Example of Bernoulli distribution.

Properties

- If $M_X(t)$ converges around 0, $M_X^{(k)}(0) = \mathbb{E}[X^k]$, meaning the moments are exactly the coefficients of the Taylor's expansion.
- **Convergence**: $M_X(t)$ converges when X is bounded.
- If independent, $M_{X+Y} = M_X M_Y$.
- **Uniqueness**: If $M_X(t)$ converges around 0, the distribution is uniquely determined by the moments. (Why? See later)

But

Moments generating function may not converge

Cauchy distribution: density function $f(x) = \frac{1}{\pi(1+x^2)}$ does not have moments for any order.

An example of non-uniqueness of moments

Log-Normal-like distribution:

density function $f_{X_n}(x) = \frac{e^{-\frac{1}{2}(\ln x)^2}}{\sqrt{2\pi x}}(1 + \alpha \sin(2n\pi \ln x))$.

k-Moments $\mathbb{E}[X_n^k] = e^{k^2/2}$ for non-negative integers k.

Characteristic functions

Definition

$\varphi_X(t) \triangleq \int_{\mathbb{R}} e^{itx} dF_X(x)$ where $i = \sqrt{-1}$ and t is real.

Properties

Convergence: It always exists.
Uniqueness: It uniquely determines the distribution.
The idea of the proof.

Uniqueness of convergent moments generating functions

If the MGF converges around 0, the characteristic functions can be extended to a zone with small imaginary part and are equal along the imaginary axis.
By the unique continuation theorem of analytic complex functions, the characteristic functions are equal.

Ready to get insights

Moments

Do moments uniquely determine the distribution?
Yes, but conditionally.

Chernoff Bounds

- Why is it so good?
- Can it be improved by non-exponential functions?
- Anything to do with moments?

What's your answer?

A story of generating function

Introduced in 1730 by Abraham de Moivre, to solve the general linear recurrence problem

Wisdom: A generating function is a clothesline on which we hang up a sequence of numbers for display. -Herbert Wilf

Application to Fibonacci numbers (by courtesy of de Moivre):

$F(x) = \sum_{n=0}^{\infty} F_n x^n = x + \sum_{n=2}^{\infty} (F_{n-1} + F_{n-2}) x^n =$
$x + xF(x) + x^2 F(x)$

$\Rightarrow F(x) = \frac{x}{1-x-x^2} = \frac{1}{\sqrt{5}} \left(\frac{\psi}{x+\psi} - \frac{\phi}{x+\phi} \right) = \sum_{n=0}^{\infty} \frac{1}{\sqrt{5}} \left(\phi^n - \psi^n \right) x^n$

$\Rightarrow F_n = \frac{1}{\sqrt{5}} \left(\phi^n - \psi^n \right).$

Brief introduction to Abraham de Moivre

- May 26, 1667-Nov. 27, 1754
- A French mathematician

- de Moivre's formula
- Binet's formula
- Central limit theorem
- Stirling's formula

Legend

- Friends: Isaac Newton, Edmond Halley, and James Stirling
- Struggled for a living and lived for mathematics
- The Doctrine of Chances was prized by gamblers
  - 2nd probability textbook in history
- Predicted the exact date of his death

References

1. http://nowak.ece.wisc.edu/SLT07/lecture7.pdf
2. When Do the Moments Uniquely Identify a Distribution