Probabilistic Method and Random Graphs

Lecture 2. Generating functions and Chernoff's Bounds [1]

Xingwu Liu

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

[1] The slides are partially based on Chapter 4 of Probability and Computing.

Preface

Questions, comments, or suggestions?

Monty Hall Problem?

Do moments uniquely determine the distribution?
Yes, but conditionally...
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, $G_X(t) \triangleq \sum_{k \geq 0} t^k Pr(X = k) = E[t^X]$.
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: $G_X^{(k)}(1) = E[X(X-1)\cdots(X-k+1)]$.

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):

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

Moment generating functions

Shortcoming of probability generating functions

Only valid for non-nagetive integer random variables.

Moment generating functions

$M_X(t) \triangleq \sum_k e^{tk} Pr(X = k) = E[e^{tX}]$.
Example of Bernoulli distribution.

Properties

If $M_X(t)$ converges in a neighborhood of 0, $M_X^{(k)}(0) = E[X^k]$, meaning that the moments are exactly the coefficients of the Taylor's expansion.
If independent, $M_{X+Y} = M_X M_Y$.
Uniqueness: If X is bounded or $M_X(t)$ converges in a neighborhood of 0, the distribution is uniquely determined by the moments. (Why? The characteristic function is uniquely extendable and uniquely determines the probability)

Bad news

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

Log-Normal-like distribution: density function
$f_{X_n}(x) = \frac{e^{\frac{1}{2}(\ln x)^2}}{\sqrt{2\pi x}}(1 + \alpha \sin(2n\pi \ln x))$.
$E[X_n^h] = e^{h^2/2}$ for non-negative integers h.

Characteristic functions

Definition

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

Properties

Convergence: it always exists.
Uniqueness: Characteristic functions are one-to-one corresponding to distributions.
The idea of the proof.
Closely related to moments generating functions (implying its uniqueness)

Chernoff bounds: inequalities derived from **all** moments

Motivation

1-moment ⇒ Markov's inequality

1- and 2-moments ⇒ Chebyshev's inequality

Q: more moments ⇒ stronger inequalities?

Examples

Flip a fair coin for n trials. Let X be the number of Heads, which is around the expectation $\frac{n}{2}$. How about its concentration?

1. Union bound makes no sense.
2. By Markov's inequality:
$Pr(X - \frac{n}{2} > \sqrt{n \ln n}) < \frac{n}{n+2\sqrt{n \ln n}} \rightsquigarrow 1$.
3. By Chebyshev's inequality: $Pr(X - \frac{n}{2} > \sqrt{n \ln n}) < \frac{1}{\ln n}$.

But can we do better since all moments exist and are small? YES!

Chernoff bounds: basic form

Chernoff bounds

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

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

Remarks

1. Note that $0 < \frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} < 1$ when $\delta > 0$. The bound in 1 exponentially deceases w.r.t. μ! And so is the bound in 2.
2. Anything to do with the moments? See the proof...

Proof of the upper tail bound

For any $\lambda > 0$,
$Pr(X \geq (1+\delta)\mu) = Pr\left(e^{\lambda X} \geq e^{\lambda(1+\delta)\mu}\right) \leq \frac{\mathbf{E}\left[e^{\lambda X}\right]}{e^{\lambda(1+\delta)\mu}}$

$$\mathbf{E}\left[e^{\lambda X}\right] = \mathbf{E}\left[e^{\lambda \sum_{i=1}^{n} X_i}\right] = \mathbf{E}\left[\prod_{i=1}^{n} e^{\lambda X_i}\right] = \prod_{i=1}^{n} \mathbf{E}\left[e^{\lambda X_i}\right]$$

Let $p_i = Pr[X_i = 1]$ for each i. Then, $\mu = \mathbf{E}[X] = \sum_{i=1}^{n} p_i$. And
$\mathbf{E}\left[e^{\lambda X_i}\right] = p_i e^{\lambda \cdot 1} + (1-p_i)e^{\lambda \cdot 0} = 1 + p_i(e^{\lambda} - 1) \leq e^{p_i(e^{\lambda}-1)}$

So, $\mathbf{E}\left[e^{\lambda X}\right] \leq \prod_{i=1}^{n} e^{p_i(e^{\lambda}-1)} = e^{\sum_{i=1}^{n} p_i(e^{\lambda}-1)} = e^{(e^{\lambda}-1)\mu}$

Thus, $Pr(X \geq (1+\delta)\mu) \leq \frac{\mathbf{E}\left[e^{\lambda X}\right]}{e^{\lambda(1+\delta)\mu}} \leq \frac{e^{(e^{\lambda}-1)\mu}}{e^{\lambda(1+\delta)\mu}} = \left(\frac{e^{(e^{\lambda}-1)}}{e^{\lambda(1+\delta)}}\right)^{\mu}$.
Let $\lambda = \ln(1+\delta) > 0$, and the proof ends

Lower tail bounds and application

Lower tail bounds

Can be proved likewise.

An tentative application

Recall the coin flipping example. By the Chernoff bound,

$$Pr(X - \frac{n}{2} > \sqrt{n \ln n}) < \frac{e^{\sqrt{n \ln n}}}{\left(1 + 2\sqrt{\frac{\ln n}{n}}\right)^{\left(\frac{n}{2} + \sqrt{n \ln n}\right)}}$$

.

Hard to identify even the order.

Is there a bound that is more *friendly*?

Chernoff bounds: a simplified form

Simplified Chernoff bounds

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

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

Application to coin flipping

$Pr(X - \frac{n}{2} > \sqrt{n \ln n}) \leq n^{-\frac{2}{3}}$. This is exponentially tighter than Chebychev's inequality. Why? Thanks to all moments and mutual independence.

Proof and Remarks

Idea of the proof

1. By the fact that $\ln(1+\delta) > \frac{2\delta}{2+\delta}$ for $\delta > 0$,
$\delta - (1+\delta)\ln(1+\delta) < -\frac{\delta^2}{2+\delta}$.
2. Use calculus to show that $\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \leq e^{-\delta^2/2}$.

Remark 1

When $1 > \delta > 0$, $-\frac{\delta^2}{2+\delta} < -\frac{\delta^2}{3}$, so $Pr(X \geq (1+\delta)\mu) \leq e^{\left(-\frac{\delta^2\mu}{3}\right)}$,
and $Pr(|X-\mu| \geq \delta\mu) \leq 2e^{\left(-\frac{\delta^2\mu}{3}\right)}$.

Remark 2

The bound is simpler but looser. Generally, it is outperformed by the basic Chernoff bound. See example.

Example: random rounding

Minimum-congestion path planning

Let $G = (V, E)$ be an undirected graph. Given a set of vertex pairs $D = \{(s_i, t_i)\}_{i=1}^{m}$, find a path P_i connecting (s_i, t_i) for every i s.t. the congestion $\max_{e \in E} cong(e)$ is minimized, where $cong(e)$ means the number of paths among $\{P_i\}_{i=1}^{n}$ that contain e.

This problem is NP-hard, but we will give an approximation algorithm based on randomized rounding.
Let's begin with an integer program, relax it into a linear program, round the solution, and analyze the approximation ratio.

ILP and its relaxation

Notation

$\mathbb{P}_i$: the set of all paths connecting s_i and t_i;
f_P^i: the indicator of whether we pick path $P \in \mathbb{P}_i$ or not;
C: the congestion in the graph.

ILP

$$\begin{array}{l} \text{Min } C \\ s.t. \sum_{P\in\mathbb{P}_i} f_P^i = 1, \forall i \\ \qquad \sum_i \sum_{e\in P\in\mathbb{P}_i} f_P^i \leq C, \forall e \\ \qquad \underline{f_P^i \in \{0,1\}}, \forall i, P \end{array} \quad \Rightarrow$$

LP

$$\begin{array}{l} \text{Min } C \\ s.t. \sum_{P\in\mathbb{P}_i} f_P^i = 1, \forall i \\ \qquad \sum_i \sum_{e\in P\in\mathbb{P}_i} f_P^i \leq C, \forall e \\ \qquad \underline{f_P^i \in [0,1]}, \forall i, P \end{array}$$

Rounding a solution to the LP

For every i, randomly pick **one** path $P_i \in \mathbb{P}_i$ with probability $f_{P_i}^i$.
Use the set $\{P_i\}_{i=1}^n$ as an approximate solution to the ILP.

Approximation ratio

Notation

C: optimum congestion of the ILP.
C^*: optimum congestion of the LP. $C^* \le C$.
X_i^e: indicator of whether $e \in P_i$.
$X^e \triangleq \sum_i X_i^e$: congestion of the edge e.
$X \triangleq \max_e X^e$: the network congestion.

Objective

We hope to show that $Pr(X > (1+k)C)$ is small for a small k. By union bound, we only need to show $Pr(X^e > (1+k)C) < \frac{1}{n^3}$ for every e.

Easy facts

$\mathbf{E}[X_i^e] = \sum_{e \in P \in \mathbb{P}_i} f_P^i$.
$\mathbf{E}[X^e] = \sum_i \mathbf{E}[X_i^e] = \sum_i \sum_{e \in P \in \mathbb{P}_i} f_P^i \le C^* \le C$.

Approximation ratio

If $C = \omega(\ln n)$, k can be arbitrarily small.

Proof: For any $0 < k < 1$, $Pr(X^e > (1+k)C) \le e^{-\frac{k^2 C}{2+k}} \le \frac{1}{n^3}$.

If $C = O(\ln n)$, $k = O(\ln n)$.

Proof: $Pr(X^e > (1+k)C) \le e^{-\frac{k^2 C}{2+k}} \le e^{-\frac{k}{2}}$ for $k \ge 2$. So, $Pr(X^e > (1+k)C) \le \frac{1}{n^3}$ when $k = 6 \ln n$.

k can be improved to be $k = O\left(\frac{\ln n}{\ln \ln n}\right)$

Proof: By the basic Chernoff bounds,

$$Pr(X^e > (1+k)C) \le \left[\frac{e^k}{(1+k)^{(1+k)}}\right]^C \le \frac{e^k}{(1+k)^{(1+k)}}.$$

When $k = \Theta\left(\frac{\ln n}{\ln \ln n}\right)$, $(1+k)\ln(1+k) = \Theta(\ln n)$ and $k - (1+k)\ln(1+k) = \Theta(\ln n)$.

Remarks of the application

Remark 1

It illustrates the practical difference of various Chernoff bounds.

Remark 2

Is it a mistake to use the inaccurate expectation?
No! It's a powerful trick.
If $\mu_L \le \mu \le \mu_H$, the following bounds hold:
Upper tail: $Pr(X \geq (1+\delta)\mu_H) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu_H}$.
Lower tail: $Pr(X \leq (1-\delta)\mu_L) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu_L}$.

Chernoff bounds + Union bound: a paradigm

A high-level picture: Want to upper-bound $Pr(\text{something bad})$.

1. By Union bound, $Pr(\text{something bad}) \le \sum_{i=1}^{Large} Pr(Bad_i)$;
2. By Chernoff bounds, $Pr(Bad_i) \le minuscule$ for each i;
3. $Pr(\text{something bad}) \le Large \times minuscule = small$.

General bounds for independent sums $X_1, ...X_n$

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

Basic Chernoff bounds remain valid.

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: $P(|X - \mathbf{E}[X]| \geq t) \leq 2e^{-\frac{2t^2}{\sum_i (b_i - a_i)^2}}$.

Remarks of Hoeffding's Inequality

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

Hoeffding's Inequality

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

Idea of the proof

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

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. $\mathbf{E}[e^{\lambda Z}] \leq (1 - \theta + \theta e^{u}) e^{\theta u}$ where $\theta = \frac{-a}{b-a}$ and $u = \lambda(b - a)$.

Define $\phi(x) \triangleq -\theta x + \ln(1 - \theta + \theta e^{x})$. Then $\mathbf{E}[e^{\lambda Z}] \leq e^{\phi(u)}$.

3. Expand $\phi(u) = \phi(0) + u\phi'(0) + \frac{u^2}{2}\phi''(v), v \in [0, u]$.
4. $\phi(0) = \phi'(0) = 0, \phi''(\cdot) \leq \frac{1}{4}$

Example: Hoeffding's Inequality, Chernoff + 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

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

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

Let $a_i = (a_{i1}, a_{i2}, ...a_{im})$ be the ith row of A. Let $Z_i = \sum_j a_{ij} b_j$. By union bound, we only need to prove $Pr(|Z_i| \geq \sqrt{4m \ln n}) \leq \frac{2}{n^2}$ for each i.

Assume that there are exactly k 1s in a_i. Then Z_i is the sum of k independent r.v. over {-1,1}, each being 1 with probability $\frac{1}{2}$.

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

References

1. http://tcs.nju.edu.cn/wiki/index.php/
2. When Do the Moments Uniquely Identify a Distribution
3. http://www.cs.princeton.edu/courses/archive/fall09/cos521/Handouts/probabilityandcomputing.pdf
4. http://www.cs.cmu.edu/afs/cs/academic/class/15859-f04/www/
5. http://nowak.ece.wisc.edu/SLT07/lecture7.pdf