Probabilistic Method and Random Graphs

Lecture 2. Moments and Inequalities [1]

Xingwu Liu

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

[1] The slides are partially based on Chapters 3 and 4 of Probability and Computing.

Preface

Questions, comments, or suggestions?

Monty Hall Problem?

Review

1. Probability axioms
2. Union Bound
3. Independence
4. Conditional probability and chain rule
   - $\Pr(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} \Pr(A_i | \bigcap_{j=1}^{i-1} A_j)$
5. Random variables: expectation, linearity, Bernoulli/binomial/geometric distribution
6. Coupon collector's problem: $\mathbb{E}[X] = nH(n) \approx n \ln n$

Coupon collector's problem: fight the salesman

Expectation is too weak

Average has nothing to do with the probability of exceeding it, Guy!

Example

- Random variables Y_α with $\alpha \geq 1$
- Let $\Pr(Y_\alpha = \alpha) = \frac{1}{\alpha}$ and $\Pr(Y_\alpha = 0) = 1 - \frac{1}{\alpha}$
- $\Pr(Y_\alpha \geq 1) = \frac{1}{\alpha}$ can be arbitrarily close to 1

But, mh...
Possible to exceed so much with high probability?

An inequality for tail probability

Markov's inequality

If $X \geq 0$ and $a > 0$, $\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$.

Proof:

$$\mathbb{E}[X] = \sum_{i \geq 0} i * \Pr(X = i) \geq \sum_{i \geq a} i * \Pr(X = i)$$
$$\geq \sum_{i \geq a} a * \Pr(X = i) = a * \Pr(X \geq a).$$

Observations

- Intuitive meaning (level of your income)
- With 12 coupons, $\mathbb{E}[X] \approx 30, \Pr(X \geq 200) < 1/6$
- Loose? Tight when only expectation is known!

Conditional expectation

Definition

$$\mathbb{E}[Y|Z = z] = \sum_y y * \Pr(Y = y|Z = z)$$

Theorem

$$\mathbb{E}[Y] = \mathbb{E}_Z[\mathbb{E}_Y[Y|Z]] \triangleq \sum_z \Pr(Z = z)\mathbb{E}[Y|Z = z]$$

Proof.

$$\begin{aligned}
\sum_z \Pr(Z = z)\mathbb{E}[Y|Z = z] &= \sum_z \Pr(Z = z) \sum_y y \frac{\Pr(Y=y, Z=z)}{\Pr(Z=z)} \\
&= \sum_y y \sum_z \Pr(Y = y, Z = z) \\
&= \sum_y y \Pr(Y = y) = \mathbb{E}[Y]
\end{aligned}$$

□

Application: expected run-time of Quicksort

Via conditional expectation

- X_n: the runtime of sorting an n-sequence.
- K: the rank of the pivot.
- If $K = k$, the pivot divides the sequence into a $(k-1)$-sequence and an $(n-k)$-sequence.
- Given $K = k$, $X_n = X_{k-1} + X_{n-k} + n - 1$.
- $\mathbb{E}[X_n | K = k] = \mathbb{E}[X_{k-1}] + \mathbb{E}[X_{n-k}] + n - 1$.
- $\mathbb{E}[X_n] = \sum_{k=1}^{n} \Pr(K = k)(\mathbb{E}[X_{k-1}] + \mathbb{E}[X_{n-k}] + n - 1)$
  $= \sum_{k=1}^{n} \frac{\mathbb{E}[X_{k-1}] + \mathbb{E}[X_{n-k}]}{n} + n - 1$.
- Please verify that $\mathbb{E}[X_n] = 2n \ln n + O(n)$.

Application: expected run-time of Quicksort

Via linearity + indicators

- y_i: the i-th biggest element
- Y_{ij}: indicator for the event that y_i, y_j are compared
- $Y_{ij} = 1$ iff the first pivot in $\{y_i, y_{i+1}, ...y_j\}$ is y_i or y_j
- $\mathbb{E}[Y_{ij}] = \Pr(Y_{ij} = 1) = \frac{2}{j-i+1}$
- $X_n = \sum_{i=1}^{n} \sum_{j=i+1}^{n} Y_{ij}$
- $\mathbb{E}[X_n] = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathbb{E}[Y_{ij}]$
- It is easy to see that $\mathbb{E}[X_n] = (2n+2) \sum_{i=1}^{n} \frac{1}{i} + O(n)$

Moments of random variables

Why moments?

- *Global* features of a random variable.
- Expectation is too weak: can't distinguish Y_α

Definition

- kth moment: $\mathbb{E}[X^k]$.
- Variance: $Var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$
  Show how far the values are away from the average.
- Examples: $Var[Y_\alpha] = \alpha - 1$

- Covariance: $Cov(X, Y) \triangleq \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.
- It's zero in case of independence.

Properties of the variance

$$Var[X+Y] = Var[X] + Var[Y] + 2Cov(X,Y)$$

$Var[X+Y] = Var[X] + Var[Y]$ if X and Y are independent.

$$Var[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

$$Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Variances of some random variables

Binomial random variable with parameters n and p

- $X = \sum_{k=1}^{n} X_i$ with the X_i's independent.
- $Var[X_i] = p - p^2 = p(1-p)$.
- $Var[X] = \sum_{k=1}^{n} Var[X_i] = np(1-p)$

Geometric random variable with parameter p

Straightforward computing shows that $Var[X] = \frac{1-p}{p^2}$

Coupon collector's problem

- We know that $Var[X_i] = \frac{1-p_i}{p_i^2}$.
- $Var[X] = \sum_{k=1}^{n} Var[X_i] \leq \sum_{k=1}^{n} \frac{n^2}{(n-k+1)^2} \leq \frac{\pi^2 n^2}{6}$

A new argument against the salesman

Chebyshev's inequality

- $\Pr(|X - \mathbb{E}[X]| \geq a) \leq \frac{Var[X]}{a^2}$.
- An immediate corollary from Markov's inequality.

Coupon collector's problem

$$\Pr(X \geq 200) = \Pr(|X - \mathbb{E}[X]| \geq 170) \leq \frac{255}{170^2} < 0.01$$

A new argument against the salesman

Chebyshev's inequality

- $\Pr(|X - \mathbb{E}[X]| \geq a) \leq \frac{Var[X]}{a^2}$.
- An immediate corollary from Markov's inequality.

Coupon collector's problem

$\Pr(X \geq 200) = \Pr(|X - \mathbb{E}[X]| \geq 170) \leq \frac{255}{170^2} < 0.01$

Trump card

- By union bound, $\Pr(|X - nH_n| \geq 5nH_n) \leq \frac{1}{n^5}$.
- Hint: Consider the probability of not containing the ith coupon after $(c+1)n \ln n$ steps.

Union bound beats the others. What a surprise!

- May 16, 1821 – December 8, 1894
- A founding father of Russian mathematics

- Probability, statistics, mechanics, geometry, number theory
- Chebyshev inequality, Bertrand-Chebyshev theorem, Chebyshev polynomials, Chebyshev bias
- Aleksandr Lyapunov, Markov brothers

Chernoff bounds: inequalities of independent sum

Motivation

- 1-moment $\Rightarrow$ Markov's inequality
- 1- and 2-moments $\Rightarrow$ Chebyshev's inequality
- Q: more information $\Rightarrow$ 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?

- Union bound makes no sense
- Markov's inequality: $\Pr(X - \frac{n}{2} > \sqrt{n \ln n}) < \frac{n}{n+2\sqrt{n \ln n}} \rightsquigarrow 1$
- Chebyshev's inequality: $\Pr(X - \frac{n}{2} > \sqrt{n \ln n}) < \frac{1}{\ln n}$
- Can we do better due to independent sum? YES!

Chernoff bounds: basic form

Chernoff bounds

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

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.

Proof of the upper tail bound

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

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

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

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

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

Lower tail bound and application

Lower tail bound

Can be proved likewise.

A tentative application

Recall the coin flipping example. By the Chernoff bound,

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

Even hard to figure out 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 $\mu = \mathbb{E}[X]$,

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

Proof and Remarks

Idea of the proof

1. $\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \leq e^{-\frac{\delta^2}{2+\delta}} \Leftrightarrow \delta - (1+\delta)\ln(1+\delta) < -\frac{\delta^2}{2+\delta} \Leftarrow$ $\ln(1+\delta) > \frac{2\delta}{2+\delta}$ for $\delta > 0$.
2. Use calculus to show that $\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \leq e^{-\frac{\delta^2}{2}}$.

Remark 1

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

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

- $G = (V, E)$ is an undirected graph. $D = \{(s_i, t_i)\}_{i=1}^m \subseteq V^2$.
- Find a path P_i connecting (s_i, t_i) for every i.
- Objective: minimize the congestion $\max_{e \in E} cong(e)$, the number of the paths among $\{P_i\}_{i=1}^m$ that contain e.

This problem is NP-hard, but we will give an approximation algorithm based on randomized rounding.

- Model as an integer program
- Relax it into a linear program
- Round the solution
- Analyze the approximation ratio

ILP and its relaxation

Notation

$\mathbb{P}_i$: the set of candidate 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		**LP**
$\text{Min } C$ $s.t. \sum_{P \in \mathbb{P}_i} f_P^i = 1, \forall i$ $\sum_i \sum_{e \in P \in \mathbb{P}_i} f_P^i \le C, \forall e$ $\underline{f_P^i \in \{0,1\}}, \forall i, P$	$\Rightarrow$	$\text{Min } C$ $s.t. \sum_{P \in \mathbb{P}_i} f_P^i = 1, \forall i$ $\sum_i \sum_{e \in P \in \mathbb{P}_i} f_P^i \le C, \forall e$ $\underline{f_P^i \in [0,1]}, \forall i, P$

Round a solution to the LP

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

Notation

C: optimum congestion of the ILP.
C^*: optimum congestion of the LP. $C^* \leq C$.
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 $\Pr(X > (1+\delta)C)$ is small for a small δ.
By union bound, we only need to show $\Pr(X^e > (1+\delta)C) < \frac{1}{n^3}$ for every e.

Apply Chernoff bound to $X^e = \sum_i X_i^e$

Prove $\Pr(X^e > (1+\delta)C) < \frac{1}{n^3}$

Easy facts

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

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

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

If $C = O(\ln n)$, $\delta = \Theta(\ln n)$

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

If $C = O(\ln n)$, δ can be improved to be $\delta = \Theta\left(\frac{\ln n}{\ln \ln n}\right)$

Proof: By the basic Chernoff bounds,

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

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

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: $\Pr(X \ge (1+\delta)\mu_H) \le \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu_H}$.
- Lower tail: $\Pr(X \le (1-\delta)\mu_L) \le \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu_L}$.

Chernoff bounds + Union bound: a paradigm

A high-level picture: Want to upper-bound $\Pr(\text{something bad})$.
1. By Union bound, $\Pr(\text{something bad}) \le \sum_{i=1}^{\text{Large}} \Pr(\text{Bad}_i)$;
2. By Chernoff bounds, $\Pr(\text{Bad}_i) \le \text{minuscule}$ for each i;
3. $\Pr(\text{something bad}) \le \text{Large} \times \text{minuscule} = \text{small}$.

Why the Chernoff bound is better? Note that it's rooted at Markov's Inequality.

Can it be improved by using functions other than exponential?

References

1. http://tcs.nju.edu.cn/wiki/index.php/
2. http://www.cs.princeton.edu/courses/archive/fall09/cos521/Handouts/probabilityandcomputing.pdf
3. http://www.cs.cmu.edu/afs/cs/academic/class/15859-f04/www/