15:853: Algorithms in the Real World
Fall 2019

Lecture 13

October 17 2019

Lecturer: Rashmi Vinayak *Scribe: Mingjie Sun*

13.1 Recap: Markov Inequality

Markov's inequality is the most basic concentration bound.

Theorem 13.1 (Markov Inequality). *Let X be a* ***non-negative*** *random variable with mean μ, then:*

$$P(X \geq a) \leq \frac{\mu}{a}$$

Proof:

$$\begin{aligned}\mu &= E[X] \\ &\geq P(x \leq a) \cdot 0 + P(X \geq a) \cdot a \\ &= P(X \geq a) \cdot a\end{aligned}$$

We have:

$$P(X \leq a) \leq \frac{\mu}{a}$$

□

One thing to note about the proof of Markov Inequality is that we are only making use of the definition of expectation here.
Further, we can replace a with $k\mu$, then Markov Inequality becomes:

$$P(X \geq k\mu) \leq \frac{1}{k}$$

13.2 Overview

In this lecture, we covered the following topics:

1. Chebyshev Inequality
2. Central limit theorem
3. Chernoff Bounds
4. Hoeffding Bounds
5. Practice Problems: balls and bins load balancing

13.3 Chebyshev Inequality

Theorem 13.2 (Chebyshev Inequality). *Let X be a random variable with mean μ and variance σ^2, then:*

$$P(|X-\mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

Proof: Recall that:

$$Var(x) = E[(X-\mu)^2]$$

Then we use Markov Inequality on random variable $(X-\mu)^2$:

$$\begin{aligned} P(|X-\mu| \geq \epsilon) &= P(|X-\mu|^2 \geq \epsilon^2) \\ &\leq \frac{E[(X-\mu)^2]}{\epsilon^2} \\ &= \frac{\sigma^2}{\epsilon^2} \end{aligned}$$

□

We can choose $\epsilon = k\sigma$ where σ is the standard deviation. Then we have:

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Compared to Markov Inequality, Chebyshev Inequality is stronger because it uses second moment/variance information.

13.4 Central Limit Theorem (CLT)

Central limit theorem is useful to give a probability bound the sum of random variables.

Before we introduce central theorem inequality, we revisit two properties on the sum/product of random variables:

1. Linearity of Expectation: For n random variables $X_1, \cdots, X_n$ (can be dependent), then:

$$E[\Sigma_{i=1}^n X_i] = \Sigma_{i=1}^n E[X_i]$$

2. For n **independent** random variables $X_1, \cdots, X_n$, we have:

$$E[\pi_{i=1}^n X_i] = \pi_{i=1}^n E[X_i]$$

Theorem 13.3 (Central Theorem Inequality). *For n **iid** random variables $X_1, \cdots, X_n$ with mean μ and variance σ^2, we have: (here $\sim$ means "distributed as")*

$$\lim_{n\to\infty} \frac{\Sigma_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \sim N(0,1)$$

13.5 Chernoff Bound

Theorem 13.4 (Chernoff Bound). *For any random variable X and any $t > 0$, we have:*

$$P(X \geq a) \leq \frac{E[e^{tX}]}{e^{ta}}$$

We can minimize the right-hand side over all t and get the tightest upper bound:

$$P(X \geq a) \leq \min_{t>0} \frac{E[e^{tX}]}{e^{ta}}$$

Proof: We apply Markov Inequality on the random variable e^{tX}:

$$\begin{aligned} P(X \geq a) &= P(e^{tX} \geq e^{ta}) \\ &\leq \frac{E[e^{tX}]}{e^{ta}} \end{aligned} \tag{13.1}$$

□

Theorem 13.5 (Chernoff bound for Binomial). *Consider the sum of n **independent** Bernoulli random variables: $X = \sum_{i=1}^{n} X_i$, where $X_i = \text{Bernoulli}(p)$. (Note that $E[X] = np$.) Then for any $\delta > 0$:*

$$P(X - np \geq \delta) \leq e^{-\frac{2\delta^2}{n}}$$
$$P(X - np \leq -\delta) \leq e^{-\frac{2\delta^2}{n}}$$

Here we assume that all n bernoulli random variables have the same distribution. However, we can also have a more general form of Chernoff bound for Binomials:

Theorem 13.6. *Consider the sum of n bernoulli random variables $X = \sum_{i=1}^{n} X_i$, where $X_i = \text{Bernoulli}(p_i)$. The mean of X is $\mu = E[X] = \sum_{i=1}^{n} p_i$. Then for all $\delta > 0$:*

$$P(X \geq (1+\delta)\mu) \leq C_\delta^\mu$$

where $C_\delta = \frac{e^\delta}{(1+\delta)^{(1+\delta)}}$

Note: In Theorem 13.5, δ is a term of **additive errors**, while in Theorem 13.6, δ is a **multiplicative error** term.

13.6 Chernoff/Hoeffding Bounds

Now we consider a generalization of Chernoff bound: Hoeffding bound.

Theorem 13.7 (Hoeffding Bound). *$X_i, \cdots X_n$ are independent random variables taking values in $[0,1]$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$. Then we have:*

$$P(X > \mu + \delta) \le e^{-\frac{\lambda^2}{2\mu+\delta}}$$

$$P(X < \mu - \delta) \le e^{-\frac{\lambda^2}{3\mu}}$$

We can let $\lambda = c\mu$, then the hoeffding bound becomes:

$$P(X > \mu + c\mu) \le e^{-\frac{c^2\mu^2}{2\mu+c\mu}} \approx e^{-c\mu} = e^{-\lambda}$$

This means that hoeffding bound gives **exponential decay**, which is much stronger than Markov and Chebyshev Inequality.

13.7 Application

13.7.1 Sampling and Opinion Polls

Problem: Suppose there is n arbitrary binary numbers in $\{0,1\}$, now we pick s of them randomly **with replacement**. Show that the sample mean is within $(1 \pm \epsilon)$ of the true mean with probability at least $1 - \delta$ if

$$s \ge \Omega\left(\frac{1}{\epsilon^2}\log\frac{1}{\delta}\right)$$

Proof: Let X_i be the random variable representing the i^{th} sampled value. Denote the true mean as μ. Then we can use Chernoff bound for bernoulli random variables on the sum $X = \sum_{i=1}^{s} X_i$:

$$\begin{aligned} P(|X - n\mu| \ge s\epsilon) &= P(|X - s\mu| \ge s\epsilon) \\ &\le 2 \cdot e^{-\frac{2(s\epsilon)^2}{s}} \\ &= 2 \cdot e^{-2s\epsilon^2} \end{aligned} \tag{13.2}$$

Now if $s \ge \frac{1}{2\cdot\epsilon^2}(\ln 2 + \log\frac{1}{\delta}) = \Omega\left(\frac{1}{\epsilon^2}\log\frac{1}{\delta}\right)$, then:

$$\begin{aligned} P(|X - s\mu| \ge s\epsilon) &\le 2 \cdot e^{-2s\epsilon^2} \\ &\le \delta \end{aligned} \tag{13.3}$$

Then:

$$P(|\frac{X}{s} - \mu| < \epsilon) \ge 1 - \delta$$

□

13.7.2 Load Balancing

Model: We have N balls and N bins, now we randomly put N balls into bins.
Question: What is the expected number of balls in each bin? **Answer**: 1.

In load balancing, we are mainly concerned with the number of balls in the **maximally loaded** bin.

Theorem 13.8. *The max-loaded bin has* $O(\frac{\log N}{\log\log N})$ *balls with probability at least* $1 - \frac{1}{N}$.

Proof: We first consider the probability of **one particular** bin receiving more than $O(\frac{\log N}{\log\log N})$ balls.
Recall **Union bound**: Suppose $A_1 \cdots A_n$ are n events, we have:

$$P(A_1 \cup A_2 \cdots \cup A_n) \leq \sum_{i=1}^{n} P(A_i)$$

Now we use Union Bound to get an upper bound on the probability of a particular bin receiving more than $O(\frac{\log N}{\log\log N})$ balls: ($\{t_1, \cdots, t_k\}$ represents a set of k distinct balls and there are $\binom{N}{k}$ such sets)

$$\begin{aligned} P(\texttt{bin i has at least k balls}) &\leq \sum_{\{t_1,\cdots,t_k\}} (\texttt{balls } t_1, \cdots t_k \texttt{ are all in bin i}) \\ &= \binom{N}{k} \cdot \left(\frac{1}{N}\right)^k \\ &= \frac{N!}{(N-k)!k!} \cdot \frac{1}{N^k} \\ &\leq \frac{N^k}{k!} \cdot \frac{1}{N^k} \\ &= \frac{1}{k!} \end{aligned} \tag{13.4}$$

Recall **Sterling's approximation**:

$$k! \sim \sqrt{2\pi k}\left(\frac{k}{e}\right)^k$$

Now we can choose $k = O(\frac{\log N}{\log\log N})$ to get the upper bound $\frac{1}{N^2}$.
In the second step, we consider the probability of there being **at least** one bin with

$O(\frac{\log N}{\log \log N})$ balls. Similarly, we use **Union bound** to get an upper bound:

$$\begin{aligned} P(\exists i,\ \texttt{bin } i \texttt{ has at least } k \texttt{ balls}) &\leq \sum_{i=1}^{n} P(\texttt{bin i has at least k balls}) \\ &\leq N \cdot \frac{1}{N^2} \\ &= \frac{1}{N} \end{aligned} \tag{13.5}$$

Therefore, the probability of the max-loaded bin having $O(\frac{\log N}{\log \log N})$ balls is at least $1 - \frac{1}{N}$. $\square$