CSE 521: Design and Analysis of Algorithms I **Spring 2016**

Lecture 2: Concentration Bounds

Lecturer: Shayan Oveis Gharan *March 30th* *Scribe: Syuzanna Sargsyan*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Laws of large numbers imply for a sequence of i.i.d. random variables $X_1, X_2, \dots$ with mean μ, the sample average, $\frac{1}{n}(X_1 + X_2 + \cdots + X_n)$, converges to μ as n goes to infinity. Concentration bounds provide a quantitative distance between the sample average and the expectation. In this lecture we review several of these fundamental inequalities. In the next few lectures we will see applications of these inequalities in designing randomized algorithms.

Let $\mathcal{D}$ be a distribution. Suppose we want to estimate the mean $\mathbb{E}[X]$ of $\mathcal{D}$ and we only have access to independent samples of $\mathcal{D}$, $X \sim \mathcal{D}$. One way to estimate the mean is to independently draw samples $X_1, X_2, ..., X_n$ from the distribution and return the empirical mean: $\frac{1}{n}\sum_{i=1}^n X_i$. By law of large numbers the empriical mean converges to $\mathbb{E}_{[}[X]]$ as $n \to \infty$. In this lecture we will prove bounds on the number of samples one needs to obtain an estimate of the mean within ϵ-additive error.

2.1 Markov's Inequality

Markov's Inequality: For any nonnegative random variable (R.V.) X and any number k,

$$\mathbb{P}[X \geq k] \leq \frac{\mathbb{E}[X]}{k}.$$

Proof.

$$\mathbb{E}[X] = \sum_i i \cdot \mathbb{P}[X = i] \geq \sum_{i \geq k} i \cdot \mathbb{P}[X = i] \geq k \cdot \sum_{i \geq k} \mathbb{P}[X = i] \geq k \cdot \mathbb{P}[X \geq k].$$

□

For example, for $k = \frac{3}{2}\mathbb{E}[X]$, we can write

$$\mathbb{P}\left[X \geq \frac{3}{2}\mathbb{E}[X]\right] \leq \frac{\mathbb{E}[X]}{\frac{3}{2}\mathbb{E}[X]} = \frac{2}{3}. \tag{2.1}$$

Example: Suppose the average grade of CSE 521 is 2.0 (out of 4.0). Give a lower bound on the fraction of students who received a grede at most 3.0. We assume that a grade can be any real number between 0.0 and 4.0.

In this example $\mathbb{E}[X] = 2.0$. Taking $k = 3.0 = \frac{3}{2}\mathbb{E}[X]$ we get that at least 1/3 of the students received grade at most 3.0.

It turns out that if the only thing that we know about X is its expectation the Markov's inequality will be the best bound we can hope for. For a tight example consider the following scenario; assume $k \geq \mathbb{E}[X]$ and let

$$X = \begin{cases} k+\epsilon & \text{w.p. } \frac{\mathbb{E}[X]}{k} \\ 0 & \text{w.p. } 1-\frac{\mathbb{E}[X]}{k} \end{cases}$$

where ϵ is very close to 0.

Application 1. We use Markov's inequality to prove an upper bound on the number of fixed points of a random permutation. Recall that a permutation is a one to one and onto map $\sigma : \{1, 2, ..., n\} \to \{1, 2, ..., n\}$. We say i is a fixed point of σ iff $\sigma(i) = i$.

Claim 2.1. *With probability at least* $1 - 1/k$ *a uniformly random permutation* σ *has at most* k *fixed points.*

Proof. The trick is to define the right random variable and then use the Markov's inequality. Define $X_i = \mathbb{I}\{\sigma(i) = i\}$ and $X = \sum X_i$. Observe that X is the number of fixed points of σ. We can write down the expectation of X using the linearity of expectation.

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \mathbb{P}[X_i] = 1.$$

The second equality uses that fact that the expectation of an indicator random variable is equal to its probability. The last equality holds since σ is a uniform permutation, i.e. $\mathbb{P}[X_i] = \dfrac{1}{n}$. Thus, by Markov's inequality $\mathbb{P}[X \geq k] \leq 1/k$. $\square$

2.2 Chebyshev's Inequality

Recall the definition of the variance:

$$\begin{aligned} \mathrm{Var}(X) &:= \mathbb{E}[X - \mathbb{E}[X]]^2 \\ &= \mathbb{E}\left[X^2 + (\mathbb{E}[X])^2 - 2X\mathbb{E}[X]\right] \\ &= \mathbb{E}[X^2] + (\mathbb{E}[X])^2 - 2\mathbb{E}[X\mathbb{E}[X]] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \end{aligned} \tag{2.2}$$

The second and the third equalities follow from the linearity of expectaion. Note that since $(X - \mathbb{E}[X])^2$ is a nonnegative random variable,

$$\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2.$$

The standard deviation of random variable X is defined as $\sigma(X) := \sqrt{\mathrm{Var}(X)}$.

Chabishev's inequality: For any random variable X and any $\epsilon > 0$,

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq \epsilon] \leq \frac{\mathrm{Var}(X)}{\epsilon^2}$$

or equivalently for any number $k > 0$,

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq k\sigma] \leq \frac{1}{k^2}$$

We can read the above inequality as follows: For any random variable X with probability at least 90%, X is within three standard deviation of its expectation.

Proof. Let $Y := (X - \mathbb{E}[X])^2 \geq 0$. By Markov's inequality

$$\mathbb{P}[Y \geq \epsilon^2] \leq \frac{\mathbb{E}[Y]}{\epsilon^2}$$

By the definition of Y,

$$\mathbb{P}\left[(X - \mathbb{E}[X])^2 \geq \epsilon^2\right] \leq \frac{\mathrm{Var}(X)}{\epsilon^2}.$$

Or, equivalently,

$$\mathbb{P}\left[|X - \mathbb{E}[X]| \geq \epsilon\right] \leq \frac{\mathrm{Var}(X)}{\epsilon^2}.$$

□

Next, we describe two applications of Chebyshev's inequality.

Application 2. Polling. Consider a large set of individuals each voting 0 or 1 on a presidency candidate, and let p be the expectation. We see that using only $O(1/\epsilon^2)$ independent samples from the set we can estimate p within and *eps*-additive error.

Let $X_1, X_2, ..., X_n$ be the votes of n independently chosen individuals in this society. Observe that, for each i,

$$X_i = \begin{cases} 1 & \text{with probability} p \\ 0 & \text{with probability} 1 - p \end{cases}$$

Define a R.V. $X = \frac{\sum X_i}{n}$. Obviously,

$$\mathbb{E}\left[X\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[X_i\right] = p.$$

We use the Chebyshev's inequality to show that for $n = O(1/\epsilon^2)$ w.h.p. X is within an additive distance ϵ of p. To use Chebyshev's inequality, we first need to upper bound the variance of X. We use the following lemma to calculate the variance of sum of independent random variables.

Lemma 2.2. *Let $X_1, X_2, ..., X_n$ be pairwise independent random variables. This means that for any $i \neq j$, $\mathbb{E}\left[X_i X_j\right] = \mathbb{E}\left[X_i\right] \mathbb{E}\left[X_j\right]$. For $X = X_1 + \ldots X_n$, we have,*

$$\mathrm{Var}(X) = \sum i = 1^n \,\mathrm{Var}(X_i).$$

Proof. By (2.2),

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \sum_{i,j} \mathbb{E}[X_i X_j] - \sum_{i,j} \mathbb{E}[X_i]\mathbb{E}[X_j],$$

where the second equality follows by linearity of expectation. By pairwise independence property, for any $i \neq j$, $\mathbb{E}[X_i X_j] = \mathbb{E}[X_i]\mathbb{E}[X_j]$. Therefore, the above expression simplifies to,

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathbb{E}[X_i^2] - \sum_{i} (\mathbb{E}[X_i])^2 = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

□

In the polling example, we can write,

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i/n) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Recall that X_i is a Bernoulli random variable with prior p. We have, $\mathrm{Var}(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2$. Obviously, $\mathbb{E}[X_i] = p$. In addition,

$$\mathbb{E}[X_i^2] = 1^2 \cdot p + 0^2 \cdot (1-p) = p.$$

So, $\mathrm{Var}(X_i) = p - p^2 \le 1/4$ and

$$\mathrm{Var}(X) \le \frac{1}{n^2} \cdot n \cdot \frac{1}{4} = \frac{1}{4n}.$$

Now, by Chebyshev's inequality

$$\mathbb{P}[|X - p| \ge \epsilon] \le \frac{\frac{1}{4n}}{\epsilon^2} = \frac{1}{4n\epsilon^2}$$

This means that for $n = 3/\epsilon^2$, X approximates p within an additive error of ϵ with 90% probability.

Application 3. Birthday Paradox. Let $X_1, ..., X_n \in \{1, 2, ..., N\}$ chosen independently and uniformly at random. How large should n be to get a *collision*, i.e., to get $X_i = X_j$ for some $i \ne j$? We show that if $n < \sqrt{N}$ then w.h.p. there is no collision. And, if $n > C.\sqrt{n}$ then with probability at least $1/C^2$ there is a collision.

Define a R.V. $Y_{ij} = \mathbb{I}(X_i = X_j)$ and let $Y = \sum_{i,j} Y_{ij}$. Note that Y_{ij}'s are dependent random variables but they are pairwise independent. This crucial fact allows us to use Lemma 2.2 to calculate the variance of Y.

Observe that Y is an integral random variables which counts the number of collisions. So, we are interested in $\mathbb{P}[Y \ge 1]$. We start by calculating the first moment of Y.

$$\mathbb{E}[Y] = \sum_{i<j} \mathbb{E}[Y_{ij}] = \sum_{i<j} \mathbb{P}[Y_{ij}] = \frac{\binom{n}{2}}{N}.$$

By Markov's inequality

$$\mathbb{P}[Y \ge 1] \le \frac{\mathbb{E}[Y]}{1} = \frac{\binom{n}{2}}{N} \approx \frac{n^2}{2N},$$

Therefore, if $n \le \sqrt{N}$ with probability at least $1/2$ there is no collisions.

Now, let us study the case where $n \ge \sqrt{N}$. Here, we use the Chebyshev's inequality. First, observe that since Y is an integral random variable,

$$\mathbb{P}[Y = 0] \le \mathbb{P}[|Y - \mathbb{E}[Y]| \ge \mathbb{E}[Y]].$$

By Chebyshev's inequality,

$$\mathbb{P}[|Y - \mathbb{E}[Y]| \ge \mathbb{E}[Y]] \le \frac{\mathrm{Var}(Y)}{(\mathbb{E}[Y])^2}.$$

Therefore

$$\mathbb{P}[Y \ge 1] = 1 - \mathbb{P}Y = 0 \ge 1 - \frac{\mathrm{Var}(Y)}{(\mathbb{E}[Y])^2}$$

Using pairwise independence of Y_{ij}'s, we get

$$\mathrm{Var}(Y) = \sum_{i<j} \mathrm{Var}(Y_{ij}) = \sum_{i<j} \left(\frac{1}{N} - \frac{1}{N^2}\right) \le \sum_{i,j} \frac{1}{N} \le \frac{\binom{n}{2}}{N}.$$

Therefore,

$$\mathbb{P}[Y \geq 1] \geq 1 - \frac{\mathrm{Var}(Y)}{\mathbb{E}[Y]^2} \geq 1 - \frac{\binom{n}{2}/N}{(\binom{n}{2}/N)^2} = 1 - \frac{N}{\binom{n}{2}} \approx 1 - \frac{2N}{n^2}.$$

So, for $n \geq C\sqrt{N}$, there is a collision with probability at least $1 - 2/C^2$.

2.3 Chernoff Bounds

Central Limit Theorems in their general form state for a sequence i.i.d. random variables $X_1, X_2, \ldots$ with bounded mean μ and variance σ^2,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \to N(0, \sigma^2)$$

Chernoff types bound provide a quantitative bound on this convergence. Recall that Chebyshev's bound imply that the probability that a R.V. X is at distance $k\sigma$ from the mean is $1/k^2$. Roughly speaking, Chernoff types of bounds imply that for a suitable R.V. X this probability is $\exp(\Omega(k))$. We start by describing the Hoeffding's bound.

Hoeffding's Inequality: Let $X_1, ..., X_n$ be a sequence of independent variables where for each $1 \leq i \leq n$, $a_i \leq X_i \leq b_i$. Then,

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]\right| \geq \epsilon\right] \leq 2\exp\left(-\frac{2n^2\epsilon^2}{\sum(a_i - b_i)^2}\right).$$

In the polling example we had $X_i \in \{0, 1\}$ for each i, and $X_1, \ldots, X_n$ are independent random variables with $\mathbb{E}[X_i] = p$. Therefore, by the Hoeffding's inequality we get

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| \geq \epsilon\right] \leq 2\exp\left(-\frac{2n^2\epsilon^2}{n}\right) = 2\exp(-2\epsilon^2 n)$$

So, for any $\delta > 0$, $\frac{1}{n}\sum_{i=1}^{n} X_i$ is within additive error ϵ of p with probability at least $1 - \delta$ if

$$n \geq \frac{\log\frac{1}{\delta}}{\epsilon^2}.$$

Application 4. Unbiased random walk on a line. Consider a particle which does an unbiased random walk on the real line. It starts at zero and in each time step it moves one step ahead or one step back, i.e., from position i with probability $1/2$ it goes to $i + 1$ and with the remaining probability it goes to $i - 1$. We want to see how far from the origin the particle will be at time n.

We can simulate this variable b a sequence $X_1, \ldots, X_n$ of independent random variables where for each i,

$$X_i = \begin{cases} 1 & \text{with probability} 1/2 \\ -1 & \text{with probability} 1/2 \end{cases}$$

Let $X = X_1 + X_2 + \cdots + X_n$. We want to prove an upper bound on $|X|$. Since $\mathbb{E}[X] = 0$, by Hoeffding inequality,

$$\mathbb{P}\left[\left|\frac{1}{n}X - 0\right| \geq \epsilon\right] \leq 2\exp\left(-\frac{2n^2\epsilon^2}{4n}\right)$$

so if $\epsilon = \sqrt{\log(n)/n}$, with probability at least $1 - 1/n$, we have $\left|\frac{1}{n}X\right| \leq \sqrt{\log(n)/n}$, or in other words, with probability at least $1 - 1/n$, $|X| \leq \sqrt{n\log(n)}$. That is, with high probability the particle is at distance $\sqrt{n\log(n)}$ from the origin. In the next lecture, we show that the particle has distance at least $\sqrt{n}$ from the origin w.h.p..