Lecture 10 (10/04/2017): Concentration bounds

Lecturer: Shi Li *Scribe: Jiayi Xian*

Suppose 10000 coin tosses are performed. We want to estimate the probability of the event that the number of total head-ups is no less than 6000. Normal distribution can be used to help estimation this probability under the circumstances that each coin toss is fair. But what if the coins are not fairly distributed and we only know the expectation of the total head-ups? This is where Chernoff bounds can help. The scenario might seem artificial here but such setting occurs a lot in the analysis of approximation algorithm.

10.1 Chernoff Bounds

Theorem 10.1. *Let $X_1, X_2, \cdots, X_n$ be* ***independent*** *random $\{0,1\}$-variables (not necessarily identically distributed). Let $X = \sum_{i=1}^n X_i$, $\mu = \mathbb{E}[X]$*
Then $\forall \delta > 0$

$$\Pr[X \geq (1+\delta)\mu] < \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu \tag{10.1}$$

$$\Pr[X \leq (1-\delta)\mu] < \left(\frac{e^\delta}{(1-\delta)^{1-\delta}}\right)^\mu \tag{10.2}$$

Remark. *To get a rough idea how big is the value on the right hand side of equation 10.1, we can take natural logarithm from both the numerator and denominator, which becomes δ and $(1+\delta)\ln 1+\delta$, respectively. Thus when δ is close to 0, r.h.s of equation 10.1 becomes*

$$\exp\Big((\delta - (1+\delta)(\delta - \delta^2/2 + \cdots))\mu\Big) \approx e^{-\frac{\delta^2}{2}\mu}$$

Thus if n is large enough, μ will also be large and the probability of X deviate from its expectation by a multiplicative factor of $1+\delta$ becomes exponentially small.

We can apply this result to the scenario we introduced in the beginning of this lecture. In that setting, suppose the coins are fair, then $n = 10000$, $\mu = 5000$, $\delta = 0.2$. Each toss is independent from the others so we can apply Chernoff Bounds:

$$\Pr[\geq 6000 \text{ head-ups}] < e^{-\frac{\delta^2}{2}\mu} = e^{-0.2^2/2\cdot 5000} = e^{-100} = 3.72 \cdot 10^{-44}$$

which I will safely put it as "impossible to happen". The analysis and result is exactly the same if we assume coins are not fair but the 10000 coin tosses has expectation of 5000 head-ups.
In the following two sections we will show another two commonly used concentration inequalities and their proofs and we will compare their preconditions and bounds.

10.1.1 Markov's inequality

Theorem 10.2. *Let X be a random variable taking non-negative values. $\mu = \mathbb{E}[X]$. Then $\forall a \leq 1$, $\Pr[X \geq a\mu] \leq \frac{1}{a}$.*

Proof: It's a one-line proof. $\mu = \mathbb{E}[X] \geq a\mu \cdot \Pr[X \geq a\mu]$. Thus $\Pr[X \geq a\mu] \leq \frac{1}{a}$ ■

10.1.2 Chebyshev's inequality

Theorem 10.3. *Let X be a random variable, $\mu = \mathbb{E}[X]$, $\delta^2 = Var[X]$. Then $\forall a \leq 1$, $\Pr[|X-\mu| \leq a\delta] \leq \frac{1}{a^2}$.*

Proof: By definition $\mathrm{Var}[X] = \mathbb{E}(|X-\mu|^2) = \delta^2$. Then apply Markov's inequality on the positive random variable $|X-\mu|^2$ we'll get $\Pr[|X-\mu| \leq a\delta] = \Pr[|X-\mu|^2 \leq a^2\delta^2] \leq \frac{1}{a^2}$. ∎

10.1.3 Comments

Chernoff bound is a "stronger" result than Markov's inequality and Chebyshev's inequality in the sense that the probability of "bad events" can be exponentially small. The precondition is that the random variable we bound is a summation of independent bounded variables, while in the Markov's inequality and Chebyshev's inequality we don't assume anything about the random variable X other than positiveness or bounded variance. It is worth to note that independence must be assumed here to make the bound holds.

10.1.4 Proof of Chernoff's Bound

Let t be a real number. Let $p_i = \Pr[X_i = 1]$

$$\begin{aligned}
\mathbb{E}[e^{tX}] &= \mathbb{E}[\prod_{i=1}^{n} e^{tX_i}] \\
&= \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}] && \text{by independency} \\
&= \prod_{i=1}^{n} [(1-p_i)) \cdot 1 + p_i \cdot e^t] \\
&= \prod_{i=1}^{n} [1 + p_i(e^t - 1)] \\
&\leq \prod_{i=1}^{n} e^{p_i(e^t-1)} && \text{because } 1 + x \leq e^x \\
&= e^{\mu(e^t-1)}
\end{aligned}$$

By Markov's Inequality, if $t > 0$

$$\Pr[X \geq (1+\delta)\mu] = \Pr[e^{tX} \geq e^{t(1+\delta)\mu}] \leq \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mu}} \leq \frac{e^{\mu(e^t-1)}}{e^{t(1+\delta)\mu}} \tag{10.3}$$

Choose $t = \ln(1+\delta)$ (not optimal, but good enough). Then the r.h.s of equation 10.3 becomes $\left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$.
Equation 10.2 can be proved in a symmetric way.

10.1.5 General Form

Theorem 10.4. *Let $X_1, X_2, \cdots, X_n$ be* ***independent*** *random $\{0,1\}$-variables. Let $X = \sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}[X]$*
Then $\forall U \geq \mu$, $\forall L \leq \mu$ $\forall \delta > 0$

$$\Pr[X \geq (1+\delta)U] < \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{U} \tag{10.4}$$

$$\Pr[X \leq (1-\delta)L] < \left(\frac{e^{\delta}}{(1-\delta)^{1-\delta}}\right)^{L} \tag{10.5}$$

$$\tag{10.6}$$

Proof: The proof in section 10.1.4 can be directly applied here:

$$\Pr[X \geq (1+\delta)U] = \Pr[e^{tX} \geq e^{t(1+\delta)U}] \leq \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)U}} \leq \frac{e^{\mu(e^t-1)}}{e^{t(1+\delta)U}} \leq \frac{e^{U(e^t-1)}}{e^{t(1+\delta)U}}$$

The rest of the proof follows. ∎

Lemma 10.5. *If $\delta \in (0,1)$, then*

$$\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \leq e^{-\frac{\delta^2}{3}}$$
$$\frac{e^{-}\delta}{(1-\delta)^{1-\delta}} \leq e^{-\frac{\delta^2}{2}}$$

Proof: Omitted. ∎