Lecture 2: Concentration Inequalities

Prof. Eric Price *Scribe: Chutong Yang & Fanqi Yan*

NOTE: THESE NOTES HAVE NOT BEEN EDITED OR CHECKED FOR CORRECTNESS

1 Overview

In the last lecture, we gave an introduction to some randomized algorithms and possible applications.

In this lecture, we learned some concentration inequality tools that will be used later in our class.

2 Question 1

Question: Suppose you flip 1000 unbiased coins and count the number of heads. How surprised would you to see head for 500, 507, 700 times respectively?

Let's define a random variable X_i with $0 \leq i \leq 1000$ for each independent result of flip coin for once and X for the total number of heads we get.

$$X = \sum_{i=1}^{1000} X_i$$

The average and variance are calculated as following:

$$E[X] = \sum_{i=1}^{1000} E[X_i] = \frac{n}{2} = 500$$

$$Var(X) = \sum_{i=1}^{1000} Var(X_i) = n[\frac{1}{2}(1-\frac{1}{2})^2 + \frac{1}{2}(0-\frac{1}{2})^2] = \frac{n}{4} = 250,\ \text{thus } \sigma = \sqrt{250} \approx 16$$

By applying the central limit theorem, we can approximate X with Gaussian density with the same average, $\mu = 500$ and variance, $\sigma = 16$:

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}16} e^{-\frac{(x-500)^2}{2*250}}$$

Thus density at 500:

$$\frac{1}{\sqrt{2\pi}\sigma} \approx \frac{1}{40} = 2.5\%$$

Similarly, the probability that

$$Pr_{\#heads=500} \approx 2.5\%$$

$$Pr_{\#heads=507} \approx 2.5\% e^{-\frac{7^2}{500}} \approx 2.5\% e^{-\frac{1}{10}} \approx 2.5\%(1 - \frac{1}{10}) = 2.25\%$$

$$Pr_{\#heads=700} \approx 2.5\% e^{-\frac{200^2}{500}} \approx 2 \times 10^{-35}$$

3 Question 2

Suppose you have a coin with unknown bias probability $p \in [0, 1]$. How many flips to estimate p with approximation error $\pm\epsilon$?

Simplifications: (1) assume $p \leq \frac{1}{2}$, don't care about constants. (2)goal is $\frac{3}{4}$ success probability.

4 Helpful Mathematical Tools

Expectation, $\mathbb{E}[X]$, is linear.

Variance, $\mathrm{Var}(X)$, is linear if independent.

4.1 Markov's Inequality

If $X \geq 0, \forall t \geq 0$

$$\begin{aligned} &Pr[X \geq t\,\mathbb{E}[X]] \leq \frac{1}{t} \\ \iff &Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t} \\ \iff &\mathbb{E}[X] \geq tPr[X \geq t] \end{aligned}$$

4.2 Chebyshev's Inequality

$$Pr[|X - \mathbb{E}[X]| \geq t] = Pr[(X - \mathbb{E}[X])^2 \geq t^2] \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2} = \frac{Var(X)}{t^2}$$

or

$$Pr[|X - \mu| \geq t] \leq \frac{\sigma^2}{t^2}$$

Now, let's revisit Q2, if we define $\hat{p} \triangleq \frac{X}{n}$, what is n s.t. $|\hat{p} - p| \leq \epsilon$, w.p. $\frac{3}{4}$, or futhermore, $1 - \delta$?

$$\mathbb{E}[\hat{p}] = p$$

$$Pr[|\hat{p} - p| \geq \epsilon] \leq \frac{Var(\hat{p})}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2}$$

since

$$Var(\hat{p}) = Var(\frac{X}{n}) = \frac{1}{n^2}Var(X) = \frac{1}{n}Var(X_i) = \frac{1}{n}p(1-p)$$

For Q2, we want n large enough that $\frac{p(1-p)}{n\epsilon^2} = \frac{1}{4}$ or δ, then

$$n = \frac{4p(1-p)}{\epsilon^2} \text{ or } n = \frac{p(1-p)}{\delta\epsilon^2}$$

where $n = \frac{4}{\epsilon^2}$ or $n = \frac{1}{\delta\epsilon^2}$ respectively suffices.

4.3 Chernoff Bound

Basic version: if $X_1, ..., X_n \in [0, 1]$ independent, $X = \sum_{i=1}^n X_i$, $\mu = \mathbb{E}[X]$

$$Pr[X \geq \mu + t] \leq e^{\frac{-2t^2}{n}}$$

$$Pr[X \leq \mu - t] \leq e^{\frac{-2t^2}{n}}$$

(The version above is the additive Chernoff bound)

Notice the additive Chernoff bound is tight for unbiased coins.

$$\begin{aligned} Pr[|\hat{p} - p| \geq t] &= Pr[|x - pn| \geq tn] \\ &\leq 2 \cdot e^{-2\frac{(\epsilon n)^2}{n}} = 2e^{-2\epsilon^2 n} \end{aligned}$$

If we want $2e^{-2\epsilon^2 n} = \sigma$,

$$n = \frac{1}{2\epsilon^2} \ln \frac{2}{\sigma}$$

Comparing to what we get from Chebyshev's inequality, $n = \frac{p(1-p)}{\delta\epsilon^2}$, our additive Chernoff bound did not get the advantage of getting smaller when p is closer to 0 or 1 like Chebyshev's inequality. We need a better version of Chernoff Bound.

$$Pr[X \leq (1-\epsilon)\mu] \leq e^{-\frac{\epsilon^2}{2}\mu}$$

$$Pr[X \geq (1+\epsilon)\mu] \leq e^{-\frac{\epsilon^2}{2+\epsilon}\mu}$$

$$\Rightarrow \forall \epsilon < 1, Pr[|X - \mu| \geq \epsilon\mu] \leq 2e^{-\frac{\epsilon^2}{3}\mu}$$

(The version above is the multiplicative form)

An interesting question is why this multiplicative version is asymmetrical. The answer is that Gaussian tail and exponential tail are quite different, and we use Gaussian to approximate only one side of the tails.

For example, consider $X_1, \cdots X_n$ with each has probability $\frac{1}{n}$ to be 1. The expectation of $X = \sum_i X_i$ is $\mathbb{E}[x] = 1$ and the variance is $\mathrm{Var}(X) = 1$ since X_i is independent of each other. The probability,

$$Pr[X = n] = \frac{1}{n^n} = e^{-n \log n}$$

If we use the Central Limit Theorem, we have

$$Pr[X = n] \approx e^{-n^2/2}$$

The Chernoff will give us $e^{-\theta(n)}$.

Now, let's revisit Q2 using the multiplicative Chernoff bound. Suppose $\epsilon < p$,

$$\begin{aligned}
& Pr[|\lambda - pn| \geq \epsilon n] \\
=& Pr[|X - pn| \geq (\frac{\epsilon}{p} pn)] \leq e^{\frac{(\epsilon/p)^2}{3} pn} = e^{\frac{\epsilon^2 n}{3p}} \\
=& \delta \text{ if } n \geq \frac{3p}{\epsilon^2} \ln \frac{2}{\delta}
\end{aligned}$$

We can now have following result. Suppose we have $n \approx \frac{p \ln \frac{2}{\delta}}{\epsilon^2}$,

if you see $pn \approx (\frac{p}{\epsilon}^2 \ln(\frac{2}{\delta}))$ heads, we can get an estimate for $p \pm \epsilon$. $\iff$ if you see $\frac{1}{\epsilon^2} \ln \frac{2}{\delta}$ heads, we can get estimate for $(1 \pm \epsilon)p$.