CS 388R: Randomized Algorithms, Fall 2019 October 22nd, 2019

Lecture 16: Concentration Inequalities

Prof. Eric Price *Scribe: Zihang He, Sabee Grewal*

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

1 Overview

In the last lecture we learned some concentration inequalities. . . .

- Markov's inequality: if $X \geq 0$, $\mathbb{P}[X \geq t] \leq \frac{E[X]}{t}$;
- Chebyshev's inequality: Let $\mu = E[X]$, $\mathbb{P}[|X - \mu| \geq t] = \mathbb{P}[(X-\mu)^2 \geq t^2] \leq \frac{E[(X-\mu)^2]}{t^2}$;
- Higher moments: for any even k, $\mathbb{P}[(X-\mu)^k \geq t^k] \leq \frac{E[(X-\mu)^k]}{t^k}$.

In this lecture we will learn more about concentration inequalities and prove the additive form of the Chernoff bound.

2 Moment Generating Functions

2.1 Gaussian refresher

Gaussian $\mathcal{N}(0,1)$: $p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Let $X \sim \mathcal{N}(0,1)$, then

$$E[X^2] = 1, \qquad E[X^k] = \frac{k!}{2^{k/2}(k/2)!} = (\Theta(k))^{k/2}.$$

With Chebyshev's inequality, we get a bound of $\mathbb{P}[(X-\mu)^2 \geq t^2] \leq \frac{1}{t^2}$. With the inequality for the k-th moment (for even k), we get a bound of $c\frac{k^{k/2}}{t^k}$. To get the tightest bound, we want to find an even k which minimizes $\frac{k^{k/2}}{t^k}$ (where t is known). Note that as $k \to k+1$, the numerator grows by $\sim \sqrt{k}$ and the denominator grows by $\sim t$, so we want $k \approx t^2$.

2.2 Moment generating functions

The moment generating function (mgf) is defined as

$$\phi(\lambda) := E[e^{\lambda(X-\mu)}].$$

By Taylor expansion, the mgf can be written as

$$\phi(\lambda) = E[1 + \lambda(X-\mu) + \frac{\lambda^2}{2}(X-\mu)^2 + \frac{\lambda^3}{3!}(X-\mu)^3 + \cdots + \frac{\lambda^k}{k!}(X-\mu)^k].$$

Note, by the linearity of expectation, we see that $\mathbb{E}[\lambda(X-\mu)] = 0$ and that term disappears from the Taylor expansion. Additionally, as λ grows, $\phi(\lambda)$ grows by a weighted combination of all moments. For larger λ, $\phi(\lambda)$ will be weighted more towards the larger moments. We get a new bound for $\mathbb{P}[X-\mu \geq t]$ by applying Markov's inequality to the mgf:

$$\mathbb{P}[X-\mu \geq t] = \mathbb{P}[e^{\lambda(X-\mu)} \geq e^{\lambda t}] \leq \min_{\lambda \geq 0} \frac{\phi(\lambda)}{e^{\lambda t}}.$$

This bound is true for all mgfs. Now we'll consider mgfs of Gaussian random variables. First, let's consider the mgf of $X \sim \mathcal{N}(0,1)$:

$$\begin{aligned}
\phi(\lambda) &= E[e^{\lambda(X-\mu)}] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{\lambda x} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\lambda)^2}{2}} e^{\frac{\lambda^2}{2}} dx \\
&= e^{\frac{\lambda^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\lambda)^2}{2}} dx && \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\lambda)^2}{2}} \text{ is the pdf of } N(\lambda, 1)\right) \\
&= e^{\frac{\lambda^2}{2}}.
\end{aligned}$$

Now, let's consider the more general Gaussian distribution $X \sim \mathcal{N}(0, \sigma^2)$:

$$E[X^2] = \sigma^2, \qquad E[X^k] = (\Theta(k)\sigma^2)^{k/2}.$$

The mgf of $X \sim \mathcal{N}(0, \sigma^2)$ is

$$\begin{aligned}
\phi(\lambda) &= E[e^{\lambda(X-\mu)}] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} e^{\lambda x} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\lambda\sigma^2)^2}{2\sigma^2}} e^{\frac{\lambda^2\sigma^2}{2}} dx \\
&= e^{\frac{\lambda^2\sigma^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\lambda\sigma^2)^2}{2\sigma^2}} dx && \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\lambda\sigma^2)^2}{2}} \text{ is the pdf of } N(\lambda\sigma^2, \sigma^2)\right) \\
&= e^{\frac{\lambda^2\sigma^2}{2}}.
\end{aligned}$$

Applying this mgf to the bound we found above, we get:

$$\mathbb{P}[X-\mu \geq t] \leq \min_{\lambda \geq 0} \frac{\phi(\lambda)}{e^{\lambda t}} = \min_{\lambda \geq 0} e^{\frac{\lambda^2\sigma^2}{2} - \lambda t} = e^{-\frac{t^2}{2\sigma^2}}$$

$\left(\text{By completing square, } \frac{\lambda^2\sigma^2}{2} - \lambda t = \frac{1}{2}(\lambda\sigma - \frac{t}{\sigma})^2 - \frac{t^2}{2\sigma^2}\right)$

3 Subgaussian, Subexponential, and Subgamma Random Variables

3.1 Subgaussian random variables

Definition 1. *A random variable X is subgaussian with "variance proxy" (a.k.a. "parameter") σ^2 if*

1. $\forall \lambda, E[e^{\lambda(X-\mu)}] \leq e^{\frac{\lambda^2\sigma^2}{2}}$
2. $\mathbb{P}[|X-\mu| \geq t] \leq 2e^{-\frac{t^2}{2\sigma^2}}$
3. $E[|X-\mu|^k] \leq k^{\frac{k}{2}}\sigma^k$

All three definitions above are equivalent up to constant factors in σ.

Lemma 2. *Any variable X bounded in $[a, a+b]$ is subgaussian with variance proxy $(\frac{b}{2})^2$.*

Proof. You will be asked to prove this on the problem set! □

Lemma 3. *If X_1, X_2 are independent subgaussian with variance proxies σ_1^2, σ_2^2, $X_1 + X_2$ is subgaussian with variance proxy $\sigma_1^2 + \sigma_2^2$.*

Proof. Assume $\mu = 0$.

$$E[e^{\lambda(X_1+X_2)}] = E[e^{\lambda(X_1)}e^{\lambda(X_2)}] = E[e^{\lambda(X_1)}]E[e^{\lambda(X_2)}] \leq e^{\frac{\lambda^2\sigma_1^2}{2}}e^{\frac{\lambda^2\sigma_2^2}{2}} = e^{\frac{\lambda^2(\sigma_1^2+\sigma_2^2)}{2}}$$

□

Note, $X \in$ subgaussian$(\sigma^2) :=$ X is subgaussian with variance proxy σ^2.

Let's consider coin flip example: sum of n coin flips $x_i \in \{0,1\}$.$X_i \in$ subgaussian$(\frac{1}{4})$, so $\sum X_i \in$ subgaussian $(\frac{n}{4})$:

$$\mathbb{P}[\sum X_i \geq \mu + t] \leq e^{-\frac{2t^2}{n}}.$$

This gives us the additive Chernoff bound!

3.2 Subexponential random variables

$Z_i = \#$ flips until heads, then

$$\mathbb{P}[Z_i = t] = \frac{1}{2^t}, \qquad \mathbb{E}[Z_i] = 2.$$

Question: how do we bound $\mathbb{P}[\sum_{i=1}^n Z_i \geq 2n + t]$?

First, we find the mgf of Z_i: $\phi(\lambda) = E[e^{\lambda(Z-2)}] = \sum_{i=1}^\infty \frac{e^{\lambda i}}{2^i}e^{-2\lambda}$

Definition 4. *A random X is subexponential with "parameter" σ^2 if*

1. $\forall\lambda$ *s.t.* $|\lambda| \le \frac{1}{\sigma}, \qquad E[e^{\lambda(X-\mu)}] \le e^{\frac{\lambda^2\sigma^2}{2}}$
2. $\mathbb{P}[|X-\mu| \ge t] \le 2e^{-\frac{t}{2\sigma}}$
3. $E[|X-\mu|^k] \le k^k\sigma^k$

Example of subexponentials:

- $\mathbb{P}[i] = \frac{1}{2^i}$
- $p(x) = e^{-x}\forall x \ge 0$
- X^2 if X is subgaussian.

Definition 5. *A random X is subgamma with "parameter" (σ^2, c) if*

1. $\forall\lambda$ *s.t.* $|\lambda| \le \frac{1}{c}, \qquad E[e^{\lambda(X-\mu)}] \le e^{\frac{\lambda^2\sigma^2}{2}}$
2. $\mathbb{P}[|X-\mu| \ge t] \le 2\ \max(e^{-\frac{t^2}{2\sigma^2}}, e^{-\frac{t}{2\sigma}})$

Next lecture we"ll continue to explore subgamma and subexponential random variables.