CS 388R: Randomized Algorithms Fall 2017

Lecture 15 — October 26, 2017

Prof. Eric Price *Scribe: Kiyeon Jeon, Xiong Zheng*

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

1 Overview

In the last lecture we intoruced Johnson Lindenstrauss lemma and covered subgaussian random variable.

In this lecture we will cover subexponential and subgamma random variables.

2 Revisit Coupon Collector problem

Let's recall the coupon collector problem where we collect coupons until we find all n types of coupons. Each one that arrives is uniform on $[n]$. Let T be the total number you collect. In the lecture 5, we can know $T = nH_n + O(n)$ by Chebyshev inequality with mean $\mathbb{E}[T]$ and variance $\mathrm{Var}[T]$. Now, we will deal with better concentration inequality with subexponential and subgamma random variables.
Let T_i be the time to catch the i–th coupon. Then $T_i \sim \mathrm{geom}(1 - \frac{i-1}{n})$. $\mathbb{E}[T_i] = \frac{1}{1-\frac{i-1}{n}} = \frac{n}{n-i+1}$ $\mathbb{E}[T] = \sum \mathbb{E}[T_i] = nH_n = \Theta(n \log n)$. How can we know $\Pr(T \geq nH_n + t)$?

When $x \sim \mathrm{geom}(p)$, $\mathbb{P}[x = t] = p(1-p)^{t-1}$ $\mathbb{P}[x \geq t] = (1-p)^{t-1} \approx e^{-pt} \neq e^{-\frac{t^2}{2\sigma^2}}$ for any σ, which means x is not subgaussian variable. So, we cannot use concentration inequality from subgaussian.

3 Concentration Inequality

3.1 Subgaussian

A variable X of mean μ is subgaussian with parameter σ if

1. *MGF* – $\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\frac{\lambda^2\sigma^2}{2}}$ for all $\lambda \in \mathbb{R}$
2. *Tail* – $\mathbb{P}[|X - \mu| > t] \leq 2e^{-t^2/2\sigma^2}$ for all $t > 0$
3. *Moment* – $\mathbb{E}[|X - \mu|^k] \leq k^{k/2}\sigma^k$ for all $k > 0$

The above three items are equivalent up to constant factors in σ

3.2 Subexponential

A variable X of mean μ is subexponential with parameter σ if

1. *MGF* – $\mathbb{E}[e^{\lambda(X-\mu)}] \le e^{\frac{\lambda^2\sigma^2}{2}}$ for all $|\lambda| < 1/\sigma$
2. *Tail* – $\mathbb{P}[|X-\mu| > t] \le 2e^{-t/2\sigma}$ for all t
3. *Moment* – $\mathbb{E}[|X-\mu|^k] \le k^k\sigma^k$ for all k

The above three items are equivalent up to constant factors in σ

3.2.1 What MGF?

Suppose $p(z) = e^{-z}$ for all $z > 0$. $\mathbb{P}[Z > t] = e^{-t}$ and $\mathbb{E}[Z] = 1$ Then, MGF is

$$\begin{aligned} E[e^{\lambda(Z-1)}] &= \int_0^\infty e^{-z}e^{\lambda(z-1)}dz = \frac{e^{-\lambda}}{1-\lambda} \quad \text{for } \lambda < 1 \\ &= \frac{1-\lambda+\frac{\lambda^2}{2}-\frac{\lambda^3}{3!}+\cdots}{1-\lambda} \\ &= 1+\frac{\lambda^2}{2}+\lambda^3(\frac{1}{2}-\frac{1}{3!})+\lambda^4(\cdots)+\cdots \le e^{4\lambda^2/2} \text{ for all } |\lambda| < \frac{1}{2} \end{aligned}$$

There is a problem that the subexponential is not closed under adding independent copies.

3.3 Subgamma

A variable X of mean μ is subgamma with parameters (σ, c) if:

1. *MGF* – $\mathbb{E}[e^{\lambda(X-\mu)}] \le e^{\frac{\lambda^2\sigma^2}{2}}$ for all $|\lambda| < \frac{1}{c}$
2. *Tail* – $\mathbb{P}[|X-\mu| > t] \le 2\max\{e^{-t^2/2\sigma^2}, 2e^{-t/2c}\}$ for all t

Observation 1. *subexponential*(σ) = *subgamma*(σ, σ)

Observation 2. *subgaussian*(σ) = *subgamma*$(\sigma, 0^+)$

Tail means

$$\mathbb{P}[|X-\mu| > t] \le \begin{cases} 2e^{-t^2/2\sigma^2}, & \text{if } |t| < \sigma^2/c \\ 2e^{-t/2c}, & \text{if } |t| > \sigma^2/c \end{cases}$$

It implies that with probability $1-\delta$, $|X-\mu| \le \max\{\sigma\sqrt{2\log(2/\delta)}, 2c\log(2/\delta)\}$

Proposition 3. *If X and Y are independent subgamma random variable i.e. $X \in$ subgamma(σ, c_1) and $Y \in$ subgamma(σ, c_2), then $X+Y \in$ subgamma$(\sqrt{\sigma_1^2+\sigma_2^2}, \max(c_1, c_2))$*

Proof. for all $\lambda < \frac{1}{\max(c_1,c_2)}$, $\mathbb{E}[e^{\lambda(X+Y)}] = \mathbb{E}[e^{\lambda X}]\,\mathbb{E}[e^{\lambda Y}] \le e^{\frac{\lambda^2\sigma_1^2}{2}}e^{\frac{\lambda^2\sigma_2^2}{2}} = e^{\frac{\lambda^2}{2}(\sqrt{\sigma_1^2+\sigma_2^2})^2}$ □

3.3.1 Coupon Collector problem

Going back to the Coupon Collector problem,

$$\begin{aligned}
&T_i \sim \text{geom}(1 - \frac{i-1}{n}) \\
\Rightarrow &T_i \in \text{subexp}(\frac{n}{n-i+1}) = \text{subgamma}(\frac{n}{n-i+1}, \frac{n}{n-i+1}) \\
\Rightarrow &\sum T_i \in \text{subgamma}(\Theta(n), \Theta(n))(\because \sqrt{\sum_{i=1}^{n}(\frac{n}{n-i+1})^2} \approx \sqrt{n^2\frac{\pi^2}{6}} = \Theta(n)) \\
\Rightarrow &\mathbb{P}[T \geq nH_n + t] \leq 2 \cdot \max\{e^{-\frac{t^2}{2n^2}}, e^{-\frac{t}{2n}}\} \leq 2 \cdot e^{-\frac{t}{2n}} \\
\Rightarrow &T \leq nH_n + 2n\log(2/\delta) = O(n \log n) \text{ with probability } 1-\delta
\end{aligned}$$

3.3.2 Bounded random variable

Let $X_i \in [0,1]$ have variance σ_i^2. How about $X = \sum_{i=1}^{n} X_i$ with independent X_i's?

Lemma 4. *If random variable $X \in [0,1]$ has variance σ, then $X \in subgamma(\sigma\sqrt{2}, 2)$*

Proof. Let $Y = X - \mathbb{E}[X]$

$$\begin{aligned}
\mathbb{E}[e^{\lambda Y}] =& \sum_{k=0}^{\infty} \frac{\mathbb{E}(\lambda Y)^k}{k!} \\
=& 1 + \sum_{k=2}^{\infty} \frac{\lambda^k \, \mathbb{E}[Y^k]}{k!} \\
\overset{(a)}{\leq}& 1 + \sum_{k=2}^{\infty} \frac{\lambda^k \sigma^2}{k!} \\
\leq& 1 + \frac{\lambda^2\sigma^2}{2} \sum_{k=0}^{\infty} \lambda^k \\
=& 1 + \frac{\lambda^2\sigma^2}{2(1-\lambda)}
\end{aligned}$$

(a) is established by $|Y^k| \leq |Y^2||Y^{k-2}| \leq |Y^2|$ □

There are two methods for the concentration inequality regarding $\sum X_i$.

1. *a naive way with subgaussian*

$$\begin{aligned}
&X_i \in \text{subgaussian}(\frac{1}{2}) \\
&X \in \text{subgaussian}(\frac{\sqrt{n}}{2}) \\
&X = \mathbb{E}[X] \pm \sqrt{2n\log(2/\delta)} \text{ w.p.} 1-\delta
\end{aligned}$$

2. *a better way with subgamma*

$$
\begin{aligned}
X_i &\in \text{subgamma}(\sqrt{2\sigma_i^2}, 2) \\
X = \sum X_i &\in \text{subgamma}(\sqrt{2\sum \sigma_i^2}, 2) \\
\mathbb{P}[|X-\mu| \geq t] &\leq \max\{e^{-\frac{t^2}{4\sum \sigma_i^2}}, e^{-\frac{t}{4}}\} = \max\{e^{-\frac{t^2}{4\text{Var}(X)}}, e^{-\frac{t}{4}}\} \\
X &= \mathbb{E}[X] \pm (2\sqrt{\text{Var}(X)\log(2/\delta)} + 4\log(2/\delta)) \text{ w.p } 1-\delta
\end{aligned}
$$

4 Proof of Distributional JL

Theorem 5. *(Distributional JL) There exists distributions on matrices $A \in \mathbb{R}^{m\times d}$ such that for any x in $\mathbb{R}^d$ $\|Ax\|_2 = (1\pm\epsilon)\|x\|_2$ with probability $1-\delta$ with $m = O(\frac{1}{\epsilon^2}\log(1/\delta))$*

Proof. Let $x \in \mathbb{R}^d$ with $\|x\|_2 = 1$. We can choose A as $\mathcal{N}(0,1)^{m\times d}$ (or $\{-1,+1\}^{m\times d}$).

$$
\begin{aligned}
A_{ji} \in \text{subgaussian}(1) &\Rightarrow A_{ji}x_i \in \text{subgaussian}(|x_i|) \\
&\Rightarrow (Ax)_j \in \text{subgaussian}(\sum |x_i|^2) = \text{subgaussian}(1)
\end{aligned}
$$

Let $Z = (Ax)_j$

$$
\begin{aligned}
Z \in \textit{subgaussian}(1) &\Rightarrow \mathbb{E}[Z^2] = 1 \\
&\Rightarrow \mathbb{E}[|Z^2|^k] = \mathbb{E}[|Z|^{2k}] \leq k^k \\
&\Rightarrow E[|Z^2-\mu|^k] \leq Z^k k^k \\
&\Rightarrow Z \in \textit{subexponential}(\sigma = \Theta(1)) \\
&\Rightarrow \sum (Ax)_j^2 \in \textit{subgamma}(\sqrt{m}, 1)
\end{aligned}
$$

Thus, we can get $\frac{\sum (Ax)_j^2}{m} = 1 \pm (\sqrt{\frac{2log(2/\delta)}{m}} + \frac{log(2/\delta)}{m})$ with probability $1-\delta$ if $m \geq \frac{\Theta(1)log(2/\delta)}{\epsilon^2}$ □