# Lecture 5 — September 8, 2016

*Prof. Eric Price* *Scribe: Garrett Goble, Ziyang Tang*

## 1 Overview

In the last lecture we took a more in depth look at Chernoff Bounds and introduced subgaussian and subexponential variables.

In this lecture we will continue talking about subgaussian variables and related random variables – subexponential and subgamma, and finally we will give a proof of famous Johnson-Lindenstrauss lemma using property of subgaussian/subgamma variables.

## 2 Review of Subgaussian

**Definition 1.** *A random variable $X$ is **subgaussian** with $\mu = E(X)$ and parameter $\sigma$ if it satisfies any of the following 3 properties:*

1. *MGF:*
$$E[e^{\lambda(X-\mu)}] \leq e^{\frac{\lambda^2\sigma^2}{2}}, \qquad \forall\lambda$$
2. *Tail:*
$$Pr[|X-\mu| > t] \leq 2e^{-\frac{t^2}{2\sigma^2}}, \qquad \forall t > 0$$
3. *Moments:*
$$E[|X-\mu|^k] \leq \sigma^k k^{k/2}, \qquad \forall k > 0$$

*The above 3 properties are equivalent with constant factor.*

## 3 Example of Coin flip

Here is an example of coin flip, we would like to use the example of its tail behavior to argue that coin flip is not a subgaussian. After that we will introduce subexponential and subgamma.

**Example of Coin Flip** : Let $Y_i$ be the number of times to flip a coin until the head is up. We could see that
$$Pr[Y_i = j] = \frac{1}{2^j}, \forall j \geq 1$$

Let $Z$ be the number of times we flip a coin until we meet with K heads. Or we could think $Z$ as the sum of individual $Y_i$.i.e.

$$Z = \sum_{i=1}^{K} Y_i$$

We can easily calculate that $E[Y_i] = 2$, and $E[Z] = 2K$, but we are interested in what $Z$ look like.

**Question:** If $K = 10^6$, what will Z look like?

According to the law of large numbers, if we let K to be large, let say $K = 10^6$, $Z$ would converge like a Gaussian with high probability to take a value from

$$2 \cdot 10^6 \pm O(10^3)$$

**Suppose $Z$ is (sub)Gaussian** :

Using the tail bound we have:

$$Pr[Z \geq 2K + t\sqrt{K}] \leq e^{-\Omega(t^2)}$$

If we set $t = \sqrt{K}$, then we have:

$$Pr[Z \geq 3K] \leq e^{-\Omega(K)}$$

If we set $t = K\sqrt{K}$, then we have:

$$Pr[Z \geq K^2 + 2K] \leq e^{-\Omega(K^3)}$$

But back to the definition of $Z$, we have:

$$Pr[Z \geq t] \geq Pr[Y_1 \geq t] = \frac{2}{2^t} = e^{-\Omega(t)}$$

The above two inequality would cast a contradiction! Thus we can see that $Z$ is not subgaussian, we need a new kinds of random variable to define it.

## 4 Subexponential and Subgamma

We would give two new definition to random variable X if it satisfies any of the following properties.

**Definition 2.** *A random variable X is **subexponential** with $\mu = E(X)$ and parameter $\sigma$ if it satisfies any of the following 3 properties:*

1. *MGF:*

$$E[e^{\lambda(X-\mu)}] \leq e^{\frac{\lambda^2\sigma^2}{2}}, \qquad \forall |\lambda| < \frac{1}{\sigma}$$

2. *Tail:*

$$Pr[|X - \mu| > t] \leq 2e^{-\frac{t}{2\sigma}}, \qquad \forall t > 0$$

3. *Moments:*

$$E[|X-\mu|^k] \le \sigma^k k^k, \qquad \forall k > 0$$

*The above 3 properties are equivalent with constant factor.*

**Example** :

Let $p(z) = e^{-z}$, $\forall z \ge 0$, $E(z) = 1$.

For the MGF $E[e^{\lambda(z-1)}]$, we have:

$$\begin{aligned}
E[e^{\lambda(z-1)}] &= \int e^{-z} e^{\lambda(z-1)} dz \\
&= e^{-\lambda} \int e^{(\lambda-1)z} dz \\
&= \frac{e^{-\lambda}}{1-\lambda} \\
&= \frac{1-\lambda+\frac{\lambda^2}{2}-\frac{\lambda^3}{3!}+...}{1-\lambda} \\
&= 1+\frac{\lambda^2}{2}+\lambda^3(\frac{1}{2}-\frac{1}{3})+... \\
&\le e^{\frac{\lambda^2}{2}}, \lambda < \frac{1}{2}
\end{aligned}$$

We can see that z here is subexponential.

Now we introduce a more general distribution – subgamma, which has a subgaussian center and a subexponential tail. Here's the definition.

**Definition 3.** *A random variable $X$ is* ***subgamma*** *with $\mu = E(X)$ and parameter $(\sigma, c)$ if it satisfies any of the following 2 properties:*

1. *MGF:*

$$E[e^{\lambda(X-\mu)}] \le e^{\frac{\lambda^2\sigma^2}{2}}, \qquad \forall|\lambda| < \frac{1}{c}$$

2. *Tail:*

$$Pr[|X-\mu| > t] \le 2max(e^{-\frac{t^2}{2\sigma^2}, -\frac{t}{2c}}), \qquad \forall t > 0$$

The above 2 properties are equivalent with constant factor and we will give a short proof here.

*Proof.* From 1 to 2.

$$Pr[X-\mu \ge t] \le \frac{E[e^{\lambda(X-\mu)}]}{e^{\lambda t}} \le e^{\frac{\lambda^2\sigma^2}{2}-\lambda t}, \forall 0 < \lambda < \frac{1}{c}$$

Since

$$\frac{\lambda^2\sigma^2}{2} - \lambda t = \frac{1}{2}(\lambda\sigma - \frac{t}{\sigma})^2 - \frac{t^2}{2\sigma^2}$$

We can divide into 2 cases and finally get that

$$Pr[X - \mu \geq t] \leq e^{\frac{\lambda^2\sigma^2}{2} - \lambda t} \leq \begin{cases} e^{-\frac{t^2}{2\sigma^2}} & if \frac{t}{\sigma^2} < \frac{1}{c} \\ e^{-\frac{t}{2c}} & otherwise \end{cases}$$

Or

$$Pr[X - \mu \geq t] \leq max(e^{-\frac{t^2}{2\sigma^2}}, e^{-\frac{t}{2c}})$$

□

**Note:** This implies with probability $1 - \delta$

$$X \leq \mu + \sigma\sqrt{2\log\frac{1}{\delta}} + C\log\frac{1}{\delta}$$

According to the definition of subgamma, we can see that

$$sugexp(\sigma) = subgamma(\sigma^2, \sigma)$$

Another property of subgamma is that any sum of 2 independent subgamma is still a subgamma.

**Lemma 4.** *If $X \in subgamma(\sigma_1^2, c_1), Y \in subgamma(\sigma_2^2, c_2)$ independent, then $X+Y \in subgamma(\sigma_1^2 + \sigma_2^2, max(c_1, c_2))$*

*Proof.* Assume X,Y with 0 mean. Then

$$\begin{aligned} E[e^{\lambda(X+Y)}] &= E[e^{\lambda X} e^{\lambda Y}] \\ &= E[e^{\lambda X}]E[e^{\lambda Y}] \\ &\leq e^{\frac{\lambda^2\sigma_1^2}{2}} e^{\frac{\lambda^2\sigma_2^2}{2}}, \lambda < \frac{1}{max(c_1, c_2)} \\ &= e^{\frac{\lambda^2(\sigma_1^2+\sigma_2^2)}{2}} \end{aligned}$$

□

## 5 Back to Coin Flip

Let $Y_i$ be the number of times to flip a coin until the head is up. We have

$$Pr[Y_i > t] \leq \frac{2}{2^t}$$

So $Y_i \in subexp(O(1)) = subgamma(O(1), O(1))$

Since the sum of subgamma is still subgamma, we will have $Z = \sum_{i=1}^{K} Y_i \in subgamma(O(K), O(1))$.

Using the tail bound of subgamma, we have:

$$Pr[z \geq 2K + t] \leq e^{-\frac{1}{2}min(\frac{t^2}{K}, t)}$$

Here we can see that Z is approximately a Gaussian with $\sqrt{K}$ deviations.

## 6 Distinct element problem

We had $Pr[Y > t] = (1-t)^n \leq e^{-tn}$ therefore Y is $subexponential(\Theta(\frac{1}{n})) = subgamma(\frac{1}{n^2}, \frac{1}{n})$.

$$\sum_{i=1}^{m} Y_i = subgamma(\frac{m}{n^2}, \frac{1}{n})$$

$$\hat{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i = subgamma(\frac{1}{n^2}, \frac{1}{nm})$$

$$Pr[|\hat{Y} - \mu| \geq \frac{\epsilon}{n}] \leq 2e^{-min(\frac{\epsilon^2 m}{2}, \frac{\epsilon m}{2})} = 2e^{-\frac{\epsilon^2 m}{2}}$$

Therefore $m = O(\frac{1}{\epsilon^2}\log\frac{1}{\delta})$ for $\delta$ failure. Therefore we do not need to split all of our elements into separate buckets.

**Remark:** If $X \in subgaussian(\sigma^2)$ and $Y = X^2 \in subexponential(O(\sigma^2))$ then

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

$$Pr[Y \geq t] \leq 2e^{-\frac{t}{2}}$$

$$Pr[|Y - E[Y]| \geq t] \leq 2e^{-\frac{t-O(1)}{2}} = 2e^{-\Omega(t)}$$

## 7 Johnson-Linderstrauss Lemma

Now we move to Johnson-Linderstrauss Lemma(JL-lemma).

At first we would like to see if $X \in subgaussian(\sigma^2)$, how is $X^2$ behave.

Let $Y = X^2$, and let us assume $X$ has zero mean. According to tail bound, $Pr[|X| \geq t] \leq 2e^{-\frac{t^2}{2}}$, Or we can rewrite it as $Pr[Y \geq t] \leq 2e^{-\frac{t}{2}}$, with centralize with mean of Y which we can consider as a constant, we have $Pr[|Y - E(Y)| \geq t] \leq 2e^{-\frac{t-O(1)}{2}} \leq 2e^{-\Omega(t)}$

So far we can conclude that $Y = X^2 \in subexponential(O(\sigma^2))$

Now we will see JL-lemma.

**Lemma 5.** *(Johnson-Linderstrauss, 1984) Let $X_1, X_2, ..., X_n \in R^d$, there exist $Y_1, Y_2, ..., Y_n \in R^m$ such that*

$$\forall i, j \in [n], \|Y_i - Y_j\|_2 \in (1 \pm \epsilon)\|X_i - X_j\|_2$$

*With $m = O(\frac{\log n}{\epsilon^2})$,* ***not dependent on d!***

To prove JL-lemma, we need first to show a similar lemma:

**Lemma 6.** *((linear) Distributional JL-lemma) There exist a distribution on $A \in R^{m\times d}$ with $m = O(\frac{1}{\epsilon}\log\frac{1}{\delta})$ such that:*

$$\forall x \in R^d, \|Ax\|_2 \in (1 \pm \epsilon)\|x\|_2$$

*with $1 - \delta$ probability over $A$.*

We first show if we have Distributional JL, how we can prove JL-lemma.

*Proof.* (DJL⇒JL) Let construct $Y_i = AX_i$. We will have for some i,j

$$\|Y_i - Y_j\| = \|A(X_i - X_j)\| \in (1 \pm \epsilon)\|X_i - X_j\| \ w.p.\ 1 - \delta$$

So

$$\|Y_i - Y_j\| \in (1 \pm \epsilon)\|X_i - X_j\|, \forall i, j,\ w.p. \geq 1 - n^2\delta$$

Just set $\delta = \frac{1}{2n^2}$ we will get valid $Y_i$. □

Now we would like to see how to prove for DJL-lemma.

*Proof.* (DJL-lemma): Let A have i.i.d subgaussian(O(1)) entries with variance 1 and zero mean. For any $X \in R^d$,

$$Y_i = (AX)_i = \sum_j A_{ij}X_j$$

Then we have

$$E[Y_i^2] = E[\sum_j A_{ij}^2 X_j^2 + \sum_{j\neq k} A_{ij}A_{ik}X_jX_k] = E[\sum_j X_j^2] = \|X\|^2$$

Thus

$$E[\|AX\|_2^2] = m\|X\|^2$$

.

Since $A_{ij} \in subgaussian(O(1))$, we can see $A_{ij}X_j \in subgaussian(O(X_j^2)), Y_i = \sum_j A_{ij}X_j \in subgaussian(O(\|X\|_2^2))$

We care about $\|AX\|_2^2$, which can be rewrite as

$$|AX\|_2^2 = \|Y\|_2^2 = \sum_i Y_i^2$$

And $Y_i \in subgamma(\|X\|_2^4, \|X\|_2^2)$ as the section begins we have showed. So $\|Y\|_2^2 \in subgamma(m\|X\|_2^4, \|X\|_2^2)$. Finally we have:

$$\begin{aligned} Pr[|\|Y\|_2^2 - E[\|Y\|_2^2]| > \epsilon m\|X\|_2^2] &< 2e^{-\frac{1}{2}min(\epsilon^2 m, \epsilon m)} \\ &= 2e^{-\frac{1}{2}\epsilon^2 m} \\ &< \delta,\ if\ m > O(\frac{1}{\epsilon^2}\log\frac{1}{\delta}) \end{aligned}$$

□

Just this week, Larsen and Nelson showed that $O(\frac{1}{\epsilon^2}\log n)$ is optimal!