MTH 995-003: Intro to CS and Big Data Spring 2014

Lecture 11 — 11 February, 2014

Inst. Mark Iwen *Scribe: Oleksii Karpenko*

1 Overview

In the last lecture we discussed Gaussian random variables and used their properties to bound the probability in Lemma 1 (see Lecture 10):

$$\mathbb{P}\left[\left|\langle\vec{g},\vec{x}\rangle^2 - \|\vec{x}\|_2^2\right| \geq t\|\vec{x}\|_2^2\right] \leq e^{-ct^2} \tag{1}$$

Recall that inequality (1) is directly related to LSH functions, where we use it to hash vectors to their length (modulo w). This time will give answers to the following questions: what if $\vec{g}$ is a zero-mean vector with independent and identically distributed entries that are not Gaussians? Suppose we are given independent g_j's with $\mathbb{E}[g_j] = 0$ and $\mathbb{V}\mathbf{ar}[g_j] = 1$. Consider

$$\mathbb{P}\left[\left|\left(\sum_{j=1}^{m} g_j \cdot x_j\right)^2 - \|\vec{x}\|_2^2\right| \geq t\|\vec{x}\|_2^2\right] \tag{2}$$

Can we still bound the probability (2) in this case? We will be working towards answers to this question over the next couple lectures.

Theorem 1 (Cramer's Theorem)**.** *Let X_1, X_2, $\ldots X_m$ be independent and identically distributed real-valued random variables with Cumulant Generating Functions (CGF) (see, e.g., [1]) defined as*

$$C_{X_l}(\Theta) := \ln\left(\mathbb{E}\left[e^{\Theta X_l}\right]\right), \qquad l \in [m]. \tag{3}$$

Then for $\forall t > 0$, the following inequality holds

$$\mathbb{P}\left[\sum_{l=1}^{m} X_l \geq t\right] \leq \exp\left\{\inf_{\Theta>0}\left(-\Theta t + \sum_{l=1}^{m} C_{X_l}(\Theta)\right)\right\} \tag{4}$$

Proof:

Let us first re-express the probability of the event by exponentiating:

$$\mathbb{P}\left[\sum_{l=1}^{m} X_l \geq t\right] = \mathbb{P}\left[\exp\left\{\Theta\sum_{l=1}^{m} X_l\right\} \geq e^{\Theta t}\right] \tag{5}$$

Applying Markov's inequality and using independence of the random variables, we have:

$$\mathbb{P}\left[\exp\left\{\Theta\sum_{l=1}^{m} X_l\right\} \geq e^{\Theta t}\right] \leq e^{-\Theta t} \cdot \mathbb{E}\left[\exp\left\{\Theta\sum_{l=1}^{m} X_l\right\}\right] \leq e^{-\Theta t} \cdot \prod_{l=1}^{m} \mathbb{E}\left[e^{\Theta X_l}\right] \tag{6}$$

Re-expressing this bound in terms of the CGF gives

$$\mathbb{P}\left[\sum_{l=1}^{m} X_l \geq t\right] \leq \exp\{-\Theta t + \sum_{l=1}^{m} C_{X_l}(\Theta)\} \tag{7}$$

Optimizing over Θ by taking the infimum in (7) completes the proof. $\square$

This theorem now allows us to prove "the real theorem" of todays lectures. Notices that we used a variant of this theorem in Lecture 7 already!

Theorem 2 (Bernstein's Inequality). *Let $X_1,\ X_2,\ \ldots X_m$ be independent, zero-mean real-valued random variables, whose moments are bounded for $n \geq 2$, such that*

$$\mathbb{E}\left[|X_l|^n\right] \leq \frac{n! R^{n-2} \sigma_l^2}{2} \qquad \forall l \in [m] \tag{8}$$

for some constants $R > 0$ and $\sigma_l > 0$. (Note that in general case, R and σ_l do not have to be equal for each X_l). Then for $\forall t > 0$ the following inequality is true

$$\mathbb{P}\left[\left|\sum_{l=1}^{m} X_l\right| \geq t\right] \leq 2 \cdot \exp\left\{\frac{-t^2/2}{\sigma^2 + Rt}\right\} \tag{9}$$

where $\sigma^2 := \sum_{l=1}^{m} \sigma_l^2$.

The above inequality tells us that the probability (9) decays quickly with the factor of t^2. Note that we can bound it by putting bounds on the moments $\mathbb{E}\left[|X_l|^n\right]$.

Proof:

Let us estimate the Moment Generating Function of X_l's using the given bounds on the $\mathbb{E}\left[|X_l|^n\right]$. The MGF can be expanded in Taylor's series:

$$\mathbb{E}\left[e^{\Theta X_l}\right] = \mathbb{E}\left[1 + \Theta X_l + \frac{\Theta^2 X_l^2}{2!} + \ldots\right] \tag{10}$$

Now using bounds on $\mathbb{E}\left[|X_l|^n\right]$, that $\mathbb{E}\left[X_l\right] = 0$, and Fubini's theorem, we have

$$\mathbb{E}\left[e^{\Theta X_l}\right] = 1 + 0 + \frac{\Theta^2}{2}\mathbb{E}\left[X_l^2\right] + \ldots = 1 + \sum_{n=2}^{\infty} \Theta^n \frac{\mathbb{E}\left[X_l^n\right]}{n!} \tag{11}$$

Using bounds on moments:

$$\mathbb{E}\left[e^{\Theta X_l}\right] \leq 1 + \frac{\sigma_l^2 \Theta^2}{2} \sum_{n=0}^{\infty} (\Theta R)^n = 1 + \frac{\sigma_l^2 \Theta^2}{2} (1 - R\Theta)^{-1} \tag{12}$$

provided that $\Theta R \in (0, 1)$ for series to converge. Finally,

$$\mathbb{E}\left[e^{\Theta X_l}\right] \leq \exp\left(\frac{\sigma_l^2 \Theta^2}{2(1 - \Theta R)}\right) \tag{13}$$

Inequality (13) bounds MGF for each X_l. This gives us what we need to use Cramer's theorem. We can now start bounding (9)

$$\mathbb{P}\left[\left|\sum_{l=1}^{m} X_l\right| \geq t\right] = \mathbb{P}\left[\sum_{l=1}^{m} X_l \geq t\right] + \mathbb{P}\left[\sum_{l=1}^{m} (-X_l) \geq t\right] \tag{14}$$

Thus,

$$\mathbb{P}\left[\left|\sum_{l=1}^{m} X_l\right| \geq t\right] \leq 2 \inf_{\Theta \in (0,1/R)} \left(\exp\left\{-\Theta t + \frac{\sigma^2 \Theta^2}{2(1-\Theta R)}\right\}\right) \tag{15}$$

from Cramer's theorem, our bounds on $\mathbb{E}\left[e^{\Theta X_l}\right]$ from (13), and our definition of σ^2. Finally, choosing $\Theta = \frac{t}{\sigma^2 + Rt} < \frac{1}{R}$ yields the desired bound. $\square$

Notice that we need moment bounds in order to use Bernstein's Inequality. The following lemma will help us get them for random variables we like.

Lemma 1.

$$\mathbb{E}\left[|X|^n\right] = n \int_0^\infty \mathbb{P}\left[|X| \geq t\right] t^{n-1} dt \qquad \forall n > 0 \tag{16}$$

Proof:

$$\int_\Omega |X|^n \, d\mathbb{P} = \int_\Omega \left(\int_0^\infty I_{\{0 \leq y \leq |X|^n\}} dy\right) d\mathbb{P} = \int_0^\infty \mathbb{P}\left[|X|^n \geq y\right] dy \tag{17}$$

by Fubini. After the change of variables $y \to t^n$, we have

$$\int_\Omega |X|^n \, d\mathbb{P} = n \int_0^\infty \mathbb{P}\left[|X|^n \geq t^n\right] t^{n-1} dt. \tag{18}$$

The desired equality follows. $\square$

We will now define the types of random variables we will care most about for the next month.

Definition 1. *We shall say that X is a sub-exponential random variable with parameters $\beta > 0$ and $\kappa > 0$ if*

$$\mathbb{P}\left[|X| \geq t\right] \leq \beta e^{-\kappa t} \qquad \forall t > 0 \tag{19}$$

Definition 2. *Similarly, we will call X a sub-gaussian random variable if $\exists\, \beta, \kappa > 0$ such that*

$$\mathbb{P}\left[|X| \geq t\right] \leq \beta e^{-\kappa t^2} \qquad \forall t > 0 \tag{20}$$

These are fairly general types of random variables. For example, any random variable X, which is bounded almost surely will be a sub-gaussian (e.g., Bernoulli, binomial, uniform on a compact set, etc.).

We can now prove another version of Bernstein's Inequality for sub-exponentials that depends only on the sub-exponential parameters β and κ.

Theorem 3 (Bernstein's Inequality for sub-exponentials). *Let X_1, X_2, $\dots X_m$ be zero-mean independent sub-exponential random variables with same parameters $\beta > 0$ and $\kappa > 0$. Then*

$$\mathbb{P}\left[\left|\sum_{l=1}^{m} X_l\right| \geq t\right] \leq 2\exp\left\{\frac{-(\kappa t)^2/2}{2\beta m + \kappa t}\right\} \qquad \forall l \in [m] \tag{21}$$

Proof:

Lemma 1 tells us that the following inequality holds for $\forall n \geq 2$

$$\mathbb{E}\left[|X_l|^n\right] = n\int_0^\infty \mathbb{P}\left[|X_l| \geq t\right] t^{n-1} dt \tag{22}$$

Using the definition of a subexponential RV, we have

$$\mathbb{E}\left[|X_l|^n\right] \leq \beta n \int_0^\infty e^{-\kappa t} t^{n-1} dt = \beta n \kappa^{-n} \int_0^\infty e^{-u} u^{n-1} du \tag{23}$$

after change of variables ($\kappa t \to u$). Note that the integral on RHS the Gamma function value $\Gamma(n) = (n-1)!$, which yields

$$\mathbb{E}\left[|X_l|^n\right] \leq \frac{\beta n!}{\kappa^n} \tag{24}$$

Applying Bernstein's inequality with $R = \frac{1}{\kappa}$, $\sigma_l^2 = \frac{2\beta}{\kappa^2}$ completes the proof. □

Next time we will work our way towards showing that sub-gaussian random matrices behave "a lot like Gaussians random matrices do". This will allow us to get concentration inequalities like Lemma 1 in lecture 10 for much more general types of random matrices.

2 Homework 3

1). Problem 7.2, p.199 of the textbook [1].
2). Problem 7.5, p.199 of the textbook [1].

References

[1] Simon Foucart and Holger Rauhut. A Mathematical Introduction to Compressive Sensing. Birkhauser Basel, 2013.