MATH 7251: High-Dimensional Probability **Spring 2022**

Lecture 6

Name: Tyler LaBonte **Professor:** Konstantin Tikhomorov

1 Bernstein Inequalities

Lemma 1.1: MGF characterization of subexponential random variables

Assume that X is a mean zero subexponential random variable. Then for every

$$-\frac{1}{c\|X\|_{\Psi_1}} \leq \lambda \leq \frac{1}{c\|X\|_{\Psi_1}} \tag{1}$$

we have

$$\mathbb{E}[\exp(\lambda X)] \leq \exp\Big(c\lambda^2 \|X\|_{\Psi_1}^2\Big). \tag{2}$$

Conversely if X is a random variable such that for some positive finite L,

$$\mathbb{E}[\exp(\lambda X)] \leq \exp\big(L^2\lambda^2\big) \tag{3}$$

for all

$$-\frac{1}{L} \leq \lambda \leq \frac{1}{L}, \tag{4}$$

then X is mean zero and

$$\|X\|_{\Psi_1} \leq cL. \tag{5}$$

Here, c is a universal constant like 10.

Proof: We prove the first part. Suppose X is a mean zero subexponential random variable. We want to bound the MGF. We can write

$$\mathbb{E}[\exp(\lambda X)] = \mathbb{E}[\exp(\lambda X) - \lambda X] \tag{6}$$

for all $\lambda \in \mathbb{R}$. We can then expand into a series as

$$\mathbb{E}\left[1 + \sum_{k=2}^{\infty} \frac{(\lambda X)^k}{k!}\right] \leq 1 + \sum_{k=2}^{\infty} \mathbb{E}\left[\frac{|\lambda X|^k}{k!}\right]. \tag{7}$$

We proved before that $\mathbb{E}\big[|X|^k\big] \leq (\tilde{c}k)^k \|X\|_{\Psi_1}^k$ for every $k \geq 1$. Applying this fact, the above is at most

$$1 + \sum_{k+2}^{\infty} \frac{|\lambda|^k}{k!} (\tilde{c}k)^k \|X\|_{\Psi_1}^k. \tag{8}$$

Using Stirling's approximation, this is at most

$$1 + \sum_{k+2}^{\infty} |\lambda|^k (c')^k \|X\|_{\Psi_1}^k . \tag{9}$$

If

$$\lambda \in \left[-\frac{1}{2c' \|X\|_{\Psi_1}}, \frac{1}{2c' \|X\|_{\Psi_1}} \right], \tag{10}$$

then for all $k \geq 2$ we have

$$|\lambda|^k (c')^k \|X\|_{\Psi_1}^k \leq 2^{2-k} \lambda^2 (c')^2 \|X\|_{\Psi_1}^2 . \tag{11}$$

Using geometric series, this is at most

$$1 + 2\lambda^2 (c')^2 \|X\|_{\Psi_1}^2 . \tag{12}$$

Using $\exp(y) \geq 1 + y$ for $y \geq 0$, this is at most

$$\exp\left(2(c')^2 \lambda^2 \|X\|_{\Psi_1}^2 \right). \tag{13}$$

So, we can take the constant from the lemma to be $c = 2(c')^2$. ■

The main "trick" to proving Bernstein-type inequalities is to study the exponentiation of the random variable (*i.e.,* its MGF) and apply Markov's inequality, then optimize over λ.

Theorem 1.1: Bernstein Inequality

Suppose $X_1, X_2, \ldots, X_n$ are independent mean zero subexponential random variables. Then,

$$\Pr\left[\left| \sum_{i=1}^n X_i \right| \geq t \right] \leq 2 \exp\left(-c \min\left(\frac{t^2}{\sum_{i=1}^n \|X_i\|_{\Psi_1}^2}, \frac{t}{\max_{1 \leq i \leq n} \|X_i\|_{\Psi_1}} \right) \right). \tag{14}$$

Note that the subexponential variables have two-tailed behavior, with the first being the subgaussian tail and the second being the subexponential tail. The critical point is

$$t_0 = \frac{\sum_{i=1}^n \|X_i\|_{\Psi_1}^2}{\max_{1 \leq i \leq n} \|X_i\|_{\Psi_1}}. \tag{15}$$

If $t \leq t_0$ we have a subgaussian tail, and if $t \geq t_0$ we have a subexponential tail. In principle, the subgaussian tail decays much faster, but notice that the scaling factor is different. In typical setups, such as when the variables are identically distributed, we would expect the scaling factor for the subexponential tail to be much smaller (it is just the max rather than sum of squares, and it does not depend on the number of variables). We will see in the proof that the subexponential tail is created by spikes of individual variables.

Proof: First we bound the MGF using the lemma.

$$\mathbb{E}[\exp(\lambda X_i)] \leq \exp\Big(c\lambda^2 \left\|X_i\right\|_{\Psi_1}^2\Big) \tag{16}$$

for

$$-\frac{1}{c \max_{1\leq i\leq n} \left\|X_i\right\|_{\Psi_1}} \leq \lambda \leq \frac{1}{c \max_{1\leq i\leq n} \left\|X_i\right\|_{\Psi_1}}. \tag{17}$$

Since we take the maximum, this works for all $1 \leq i \leq n$. Thus,

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^n X_i\right)\right] = \prod_{i=1}^n \mathbb{E}[\exp(\lambda X_i)] \leq \exp\left(c\lambda^2 \sum_{i=1}^n \left\|X_i\right\|_{\Psi_1}^2\right). \tag{18}$$

We now apply Markov's inequality. For all $0 \leq t \leq \infty$,

$$\Pr\left[\sum_{i=1}^n X_i > t\right] = \Pr\left[\exp\left(\lambda \sum_{i=1}^n X_i\right) \geq \exp(\lambda t)\right] \leq \frac{\mathbb{E}[\exp(\lambda \sum_{i=1}^n X_i)]}{\lambda t}. \tag{19}$$

Combining with the previous estimate, this is at most

$$\exp\left(c\lambda^2 \sum_{i=1}^n \left\|X_i\right\|_{\Psi_1}^2 - \lambda t\right). \tag{20}$$

We will now optimize over admissible λ for every $t > 0$. First consider when $t \leq t_0$. We pick

$$\lambda = \frac{t}{2c \sum_{i=1}^n \left\|X_i\right\|_{\Psi_1}^2} \tag{21}$$

This λ is in the interval we defined before. Applying the previous estimate and simplifying,

$$\Pr\left[\sum_{i=1}^n X_i \geq t\right] \leq \exp\left(-\frac{1}{4c} \frac{t^2}{\sum_{i=1}^n \left\|X_i\right\|_{\Psi_1}^2}\right). \tag{22}$$

We can put absolute values on the left-hand side and scale the right-hand side by a factor of 2. Now consider when $t \geq t_0$. We pick

$$\lambda = \frac{1}{2c \max_{1\leq i\leq n} \left\|X_i\right\|_{\Psi_1}}. \tag{23}$$

This λ is in the interval we defined before. Applying the previous estimate,

$$\Pr\left[\sum_{i=1}^n X_i \geq t\right] \leq \exp\left(\frac{1}{4c} \frac{\sum_{i=1}^n \left\|X_i\right\|_{\Psi_1}^2}{(\max_{1\leq i\leq n} \left\|X_i\right\|_{\Psi_1})^2} - \frac{t}{2c \max_{1\leq i\leq n} \left\|X_i\right\|_{\Psi_1}}\right). \tag{24}$$

Using the fact that $t \geq t_0$, this is at most

$$\exp\left(\frac{1}{4c}\frac{t}{\max_{1\leq i\leq n}\|X_i\|_{\Psi_1}} - \frac{t}{2c\max_{1\leq i\leq n}\|X_i\|_{\Psi_1}}\right) \leq \exp\left(-\frac{1}{4c}\frac{1}{\max_{1\leq i\leq n}\|X_i\|_{\Psi_1}}\right). \tag{25}$$

■

2 Corollaries of Bernstein Inequalities

Let $\boldsymbol{x} \in \mathbb{R}^n$ be a random vector. It is called k-subgaussian if all its one-dimensional orthogonal projections are k-subgaussian, that is

$$\|\langle \boldsymbol{x}, \boldsymbol{y}\rangle\|_{\Psi_2} \leq k \tag{26}$$

for all unit $\boldsymbol{y}$. The Hoeffding-type inequality discussed before implies that if $\boldsymbol{x}$ has independent mean zero subgaussian components, then

$$\|\boldsymbol{x}\|_{\Psi_2} \leq c \max_{1\leq i\leq n}\|X_i\|_{\Psi_2} \tag{27}$$

where c is a universal constant.

Corollary 2.1

If $\boldsymbol{x}$ is a subgaussian centered vector in $\mathbb{R}^n$ with independent components and $\boldsymbol{x}$ is isotropic, that is, $\mathbb{E}\left[\boldsymbol{x}\boldsymbol{x}^\top\right] = \mathbf{I}$, then

$$\Pr\left[\left|\|\boldsymbol{x}\|_2^2 - n\right| \geq t\right] \leq 2\exp\left(-c\min\left(\frac{t^2}{\sum_{i=1}^n \|x_i\|_{\Psi_2}^4}, \frac{t}{\max_{1\leq i\leq n}\|x_i\|_{\Psi_2}^2}\right)\right). \tag{28}$$

Note that we have the same two-tailed behavior here, but with different powers. This is because the Euclidean norm of $\boldsymbol{x}$ is the sum of the squares of its components, which are independent and subexponential, and we can apply the conversion from subgaussian norm to subexponential norm which adds an extra power.

Another important remark is that there is a universal constant $\tilde{c}$ such that every variable of variance 1 has subgaussian moment at least $\tilde{c}$.

Corollary 2.2

Assume that $X_1, X_2, \ldots, X_n$ are iid centered subgaussian variables of unit variance. Then,

$$\Pr\left[\left|\|(X_1, X_2, \ldots, X_n)\|_2^2 - n\right| \geq t\right] \leq 2\exp\left(-\tilde{c}\min\left(\frac{t^2}{n\|X_1\|_{\Psi_2}^4}, \frac{t}{\|X_1\|_{\Psi_2}^2}\right)\right). \tag{29}$$

If $\|X_1\|_{\Psi_2}$ is a constant, we get subgaussian behavior up to $t_0 = n$ and subexponential afterwards.

Corollary 2.3

If $G = (g_1, g_2, \ldots, g_n)$ is a standard Gaussian vector in $\mathbb{R}^n$ then

$$\Pr\left[\left|\|G\|_2^2 - n\right| \geq t\right] = 2\exp\left(-\Theta\left(\min\left(\frac{t^2}{n}, t\right)\right)\right) \tag{30}$$

for all $t \geq 10\sqrt{n}$.

Corollary 2.4

Let $\boldsymbol{x}$ be a centered isotropic random vector with independent components having unit variances. Then,

$$\Pr\left[\left|\|\boldsymbol{x}\|_2 - \sqrt{n}\right| \geq t\right] \leq 2\exp\left(-\frac{ct^2}{\max_{1\leq i\leq n}\|X_i\|_{\Psi_2}^4}\right) \tag{31}$$

for all $t \geq 0$.

Now, if $\mathbf{A}$ is an $m \times n$ random matrix with independent k-subgaussian centered entries of unit variances. Then, for every fixed unit vector $\boldsymbol{y}$, the product $\mathbf{A}\boldsymbol{y}$ is a ck-subgaussian random vector in $\mathbb{R}^m$ with zero mean and identity covariance matrix. So, the above deviation estimates also work for matrix-vector products.