Lecture 04: Concentration Inequalities III

Nikola Zlatanov

Innopolis University

Advanced Statistics

13-th of Feb to 20-th of Feb, 2023

Motivation

- Recall that the bound on the tail of normalized sum of Gaussian RVs decays with rate which at least with $e^{-t^2/2}$.
- Recall that the bound on the tail of the sum of symmetric Bernoulli decays with rate which at least with $e^{-t^2/2}$, using Hoeffding's inequality.
- Recall that the bound on the tail of the sum of bounded RVs decays with rate which at least with e^{-ct^2}, using Hoeffding's inequality.
- Now the question is: Is this all? Are there any other RVs whose tail of the normalized sum decays also exponentially with t^2?
- More precisely: What is the biggest class of distributions for i.i.d. X_i's such that the following bound holds

$$\Pr\left\{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i - E[X_i]}{\sigma}\right| \geq t\right\} \leq c_1 e^{-c_2 t^2}, \text{ for } t \geq 0 \qquad (1)$$

Sub-Gaussian RVs

- Well, if (1) holds in general for $X_1, X_2, ..., X_n$, then it holds also for $n = 1$, in which case we obtain

$$\Pr\left\{\left|\frac{X_1 - E[X_1]}{\sigma}\right| \geq t\right\} \leq c_1 e^{-c_2 t^2}, \text{ for } t \geq 0 \tag{2}$$

- Hence, each term in the sum in (1) must themselves satisfy (2).
- The RV X that satisfies (2) is called a sub-Gaussian RV.
- The term sub-Gaussian intuitively means that the tail of X behaves like a Gaussian tail, up to some constants.
- Note, X is not Gaussian, only its tail behaves like a Gaussian tail, up to some constants.

Sub-Gaussian RVs

- Note that instead of (2), I can investigate can $\Pr\left\{|X_1 - E[X_1]| \geq \hat{t}\right\}$ since this only changes the constants $t = \sigma\hat{t} \Rightarrow \hat{t} = \frac{t}{\sigma}$

$$\begin{aligned}\Pr\{|X_1 - E[X_1]| \geq t\} &= \Pr\left\{|X_1 - E[X_1]| \geq \sigma\hat{t}\right\}\\ &= \Pr\left\{\frac{|X_1 - E[X_1]|}{\sigma} \geq \hat{t}\right\}\\ &\leq c_1 e^{-c_2\hat{t}^2} = c_1 e^{-\frac{c_2}{\sigma^2}t^2} = c_1 e^{-\hat{c}_2 t^2} \qquad (3)\end{aligned}$$

- In addition, for all sub-Gaussian RVs, without loss of generality we can assume $E[X] = 0$.
- This is only for simplicity of notations, as not to drag the terms $X_i - E[X_i]$ in the equations, since always X with $E[X] = 0$ can be obtained from RV $\hat{X}$ with $E[\hat{X}] \neq 0$ using $X = \hat{X} - E[\hat{X}]$.

Sub-Gaussian RVs

- Thm (Sub-Gaussian Lemma): For any sub-Gaussian RV X, the following statements are equivalent:
   1. Tails: there $\exists c_1 > 0$, such that the following holds
   $$\Pr\{|X| \geq t\} \leq 2e^{-t^2/c_1^2}, \text{ for } t \geq 1 \tag{4}$$
   2. Moments: there $\exists c_1 > 0$, such that the following holds
   $$(E[|X|^p])^{1/p} = c_2\sqrt{p} \tag{5}$$
   3. There $\exists c_2$, such that the following holds
   $$E[e^{X^2/c_3^2}] \leq 2 \tag{6}$$
   4. MGF: there $\exists c_3$, such that the following holds
   $$E[e^{\lambda X}] \leq e^{\lambda^2 c_4^2} \tag{7}$$

Sub-Gaussian RVs

- The sub-Gaussian lemma says that if X satisfies anyone of the properties (4) to (7), then it satisfies all properties (4) to (7).
- Note, sub-Gaussian RVs can have better tails than the Gaussian tail, but cannot have worse tails than the Gaussian tail, up to some constants.
- To prove the sub-Gaussian properties, we will use the following: Lemma (Integral Identity): $\forall$ non-negative RVs X, the following holds

$$E[X] = \int_0^\infty \Pr\{X \geq t\}\, dt \tag{8}$$

Proof: Let $\mathbf{1}(b) = 1$ if $b =$ True and $\mathbf{1}(b) = 0$ if $b =$ False. Then, $\forall x \in \mathbb{R}$,

$$x = \int_0^x 1\, dt = \int_0^\infty \mathbf{1}(t \leq x) dt \qquad \int f(x) \tag{9}$$

$E(g(x)) = \int_0^{+\infty} g(x) \cdot f_X(x) dx$

Now in (9), we substitute x with an RV X and take expectations from both sides, to obtain

$$E[X] = E\left[\int_0^\infty \mathbf{1}(t \le x)dt\right] = \int_0^\infty \left[\int_0^\infty \mathbf{1}(t \le x)dt\right] f_X(x)dx$$
$$= \int_0^\infty \left[\int_0^\infty \mathbf{1}(t \le x) f_X(x)dx\right] dt = \int_0^\infty E\left[\mathbf{1}(t \le X)\right] dt$$
$$= \int_0^\infty \Big(1 \times \Pr\{t \le X\} + 0 \times \Pr\{t \le X\}\Big) dt = \int_0^\infty \Pr\{X \ge t\}dt$$

Q.E.D. (Integral Identity)

$$\mathbf{1}(t \le X) = \begin{cases} 1, & \text{if outcome of } X \text{ is} \ge t \\ 0, & \text{otherwise} \end{cases}$$

- Proof of sub-Gaussian Lemma Property 3 using Property 2 :

$$E\left[e^{X^2/c_2^2}\right] \overset{(a)}{=} E\left[\sum_{k=0}^{\infty} \frac{X^{2k}}{c_2^{2k} k!}\right] = \sum_{k=0}^{\infty} \frac{E[X^{2k}]}{c_2^{2k} k!} \overset{(b)}{\leq} \sum_{k=0}^{\infty} \frac{c_1^{2k}(2k)^k}{c_2^{2k} k!}$$

$$\overset{(c)}{\leq} \sum_{k=0}^{\infty} \frac{c_1^{2k}(2k)^k}{c_2^{2k}(k/e)^k} = \sum_{k=0}^{\infty} \frac{c_1^{2k}(2)^k}{c_2^{2k}(1/e)^k} = \sum_{k=0}^{\infty} \left(\frac{2ec_1^2}{c_2^2}\right)^k$$

$$\overset{(d)}{=} \frac{1}{1 - \frac{2ec_1^2}{c_2^2}} \overset{(e)}{=} 2,$$

where (a) is from the Taylor series expansion of e^x, given by $e^x = \sum_{k=0}^{\infty} x^k/k!$, (b) is from (10), (c) is from Sterling's bound $k! \geq k^k/(e^k)$, (d) holds for some $c_2 > c_1\sqrt{2e}$ where we use the identity $\sum_{k=0}^{\infty} x^k = 1/(1-x)$ for $0 < x < 1$, and (e) holds if $c_2 = 2c_1\sqrt{e}$.

Sub-Gaussian RVs

- Proof of sub-Gaussian Lemma Property 4 using Property 3: Note that $(ax - b\lambda)^2 \geq 0$ always holds. Expanding, we obtain $a^2x^2 - 2abx\lambda + b^2\lambda^2 \geq 0$ always hold, or equivalently $2abx\lambda \leq a^2x^2\alpha^2 + b^2x\lambda^2$, which is equivalent to

$$x\lambda \leq \frac{\alpha}{2}x^2 + \frac{1}{2\alpha}\lambda^2, \tag{11}$$

where $\alpha = a/b$.
We can now prove our inequality for $|\lambda| > \sqrt{\ln(2)}$ as

$$\begin{aligned} E[e^{\lambda X}] &\overset{(a)}{\leq} E[e^{\frac{1}{2\alpha}\lambda^2} e^{\frac{\alpha}{2}X^2}] = e^{\frac{1}{2\alpha}\lambda^2} E[e^{\frac{\alpha}{2}X^2}] \\ &\overset{(b)}{\leq} 2e^{\frac{1}{2\alpha}\lambda^2} \overset{(c)}{=} e^{\lambda^2 c_4^2}, \text{ if } |\lambda| > \sqrt{\ln(2)} \end{aligned}$$

$e^{\lambda^2} e^{\lambda^2 \cdot \frac{1}{2\alpha}}$

where (a) is due to (11),(b) is due to Property 3 when $\alpha \leq e/(2c_2^2)$ and (c) holds if $|\lambda| > \sqrt{\ln(2)}$ and by setting $c_4^2 = 1 + 1/(2\alpha) \geq 1 + e/c_2^2$.

Sub-Gaussian RVs

- Proof of sub-Gaussian Lemma Property 4 using Property 3 when $|\lambda| \leq \sqrt{\ln(2)}$?
- **If someone can provide me with such proof, I give 5pts!**

Sub-Gaussian RVs

- Proof of sub-Gaussian Lemma Property 1 using Property 4:

$$\Pr\{X \geq t\} = \Pr\left\{e^{\lambda X} \geq e^{\lambda t}\right\} \overset{(a)}{\leq} \frac{E[e^{\lambda X}]}{e^{\lambda t}} \overset{(b)}{\leq} \frac{e^{\lambda^2 c_4^2}}{e^{\lambda t}} = e^{\lambda^2 c_4^2 - \lambda t}$$
$$\overset{(c)}{=} e^{\frac{t^2 c_4^2}{4} - \frac{t^2}{2}} \overset{(d)}{=} e^{-t^2/c_1^2},$$

where (a) is due to Markov, (b) is due to Property 4, (c) is by setting $\lambda = t/2$, and (d) is by setting $c_1^2 = 2/(2 - c_4^2)$, for $c_4^2 < 2$. Now, it is also easy to obtain from above that

$$\Pr\{-X \geq t\} \leq e^{-t^2/c_1^2},$$

also holds. Combining the talls of $\Pr\{X \geq t\}$ and $\Pr\{-X \geq t\}$, we obtain

$$\Pr\{|X| \geq t\} \leq 2e^{-t^2/c_1^2},$$

Q.E.D

Sub-Gaussian RVs

- We now can prove the following theorem
- Thm: If X_i, for $i = 1, 2, ..., n$, are independent sub-Gaussian, then $\sum_{i=1}^{n} X_i$ is also sub-Gaussian.
- Proof: We need to prove one of the four Properties, since they are all equivalent. Let's prove Property 4, i.e., that there $\exists \hat{c}_4$ (given by $\hat{c}_4^2 = \sum_{i=1}^{n} c_{4,i}^2$, for $i = 1, 2..., n$ such that

$$E\left[e^{\lambda \sum_{i=1}^{n} X_i}\right] \leq e^{\lambda^2 \sum_{i=1}^{n} c_{4,i}^2} \leq e^{\lambda^2 n \max\limits_i c_{4,i}^2} \tag{12}$$

Let's start

$$E\left[e^{\lambda \sum_{i=1}^{n} X_i}\right] = E\left[\prod_{i=1}^{n} e^{\lambda X_i}\right] \overset{(a)}{=} \prod_{i=1}^{n} E[e^{\lambda X_i}] \overset{(b)}{\leq} \prod_{i=1}^{n} e^{\lambda^2 c_{4,i}^2}$$

$$= e^{\lambda^2 \sum_{i=1}^{n} c_{4,i}^2} \overset{(c)}{\leq} e^{\lambda^2 n \max\limits_i c_{4,i}^2}$$

where (a) comes from independence, (b) comes from Property 4 for each X_i, and (c) comes from $\sum_{i=1}^{n} c_{4,i}^2 \leq n \max\limits_i c_{4,i}^2$

Sub-Gaussian RVs

- Now, since we proved that if X_i, for $i = 1, 2, ..., n$ are independent sub-Gaussian, then $\sum_{i=1}^{n} X_i$ is also sub-Gaussian, we can state Property 1 for $\sum_{i=1}^{n} X_i$ as the following theorem
- Thm (General Hoeffding's inequality): If X_i, for $i = 1, 2, ..., n$ are independent sub-Gaussian, then the following holds

$$\Pr\left\{\left|\sum_{i=1}^{n} X_i\right| \geq t\right\} \leq 2e^{-\frac{t^2}{\sum_{i=1}^{n} c_{1,i}^2}} \leq 2e^{-\frac{t^2}{n \max\limits_i c_{1,i}^2}} \tag{13}$$

- Proof: Since $\sum_{i=1}^{n} X_i$ are sub-Gaussian, Property 1 holds, from which (13) follows.
- Note that (13) is the Hoeffding's inequality but now it is more general since it holds for all sub-Gaussian RVs and now it is derived using a different method than before.

Sub-Gaussian RVs

- Examples of Sub-Gaussian
  - Gaussian
  - Any discrete or continuous distribution bounded on a finite interval.

Sub-Exponential RVs

- Although sub-Gaussian cover a wide range of distributions, there are many other important distributions that do not belong to the sub-Gaussian class of distributions.
- For example: The Poisson distribution, the Exponential distribution, etc.
- That is why we need another class of distributions.
- The class of distribution that we will investigate now, is the class of sub-exponential distributions.
- Sub-exponential distributions are especially useful when: If X is sub-Gaussian, in which case X^2 is sub-exponential. To see this, let $Y = X^2$, and let X be sub-Gaussian. Let's check the tail of Y

$$\Pr\{Y > t\} = \Pr\{X^2 > t\} = \Pr\{|X| > \sqrt{t}\} \leq 2e^{-t/c},$$

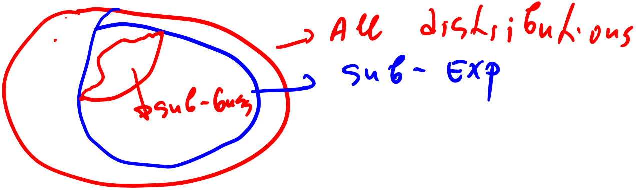
where $c > 0$ is some constant. Hence, the tail of Y is not sub-Gaussian, thereby Y is not a sub-Gaussian RV.

Sub-Exponential RVs

- Sub-exponential Lemma: For any sub-exponential RV X, the following statements are equivalent:
  1. Tails: There $\exists k_1 > 0$, such that the following holds
  $$\Pr\{|X| \geq t\} \leq 2e^{-t/k_1}, \text{ for } t \geq 1 \tag{14}$$
  2. Moments: There $\exists k_2$, such that the following holds
  $$(E[|X|^p])^{1/p} \leq k_2 p \tag{15}$$
  3. There $\exists k_3$, such that the following holds
  $$E[e^{|X|/k_3}] \leq 2 \tag{16}$$
  4. MGF: There $\exists k_4$, for X with $E[X] = 0$, such that the following holds
  $$E[e^{\lambda X}] \leq e^{\lambda^2 k_4^2}, \text{ but only for } |\lambda| < \frac{1}{k_4}, \tag{17}$$
  otherwise, for $|\lambda| \geq \frac{1}{k_4}$, $E[e^{\lambda X}] = \infty$.
- Proof: By using the same methods as per the sub-Gaussian. DIY

Sub-Exponential RVs

- Examples of Sub-exponential RVs
  - Any sub-Gaussian since its tail is $e^{-t^2} < e^{-t}$
  - Any sub-Gaussian RV squared
  - Exponential distribution: $f_X(x) = \frac{1}{\lambda} e^{-\frac{1}{\lambda}}$, $x \geq 0$.
  - Poisson distribution. The tail decays with $e^{-t\ln(t)} < e^{-t}$
- How about the tail of the sum of sub-exponentials $\sum_{i=1}^{n} X_i$, where X_i, $\forall i$ are sub-exponential RVs? Does its tail also decay with e^{-t}? Maybe it decays slower, or faster?
- Here, we have a surprise, given by the theorem on the following page.



Sub-Exponential RVs

- Thm (Bernstein's Inequality): The tail of $\sum_{i=1}^n X_i$, where X_i, $\forall i$ are sub-exponential RVs with $E[X_i]=0$, satisfies

$$
\begin{aligned}
\Pr\left\{\left|\sum_{i=1}^n X_i\right| \geq t\right\} &\leq 2\exp\left(-\min\left\{\frac{t^2}{\sum_{i=1}^n k_{4,i}^2}, \frac{t}{\max\limits_i k_{4,i}}\right\}\right)\\
&\leq 2\exp\left(-\min\left\{\frac{t^2}{n\max\limits_i k_{4,i}^2}, \frac{t}{\max\limits_i k_{4,i}}\right\}\right)\\
&= \begin{cases} 2\exp\left(-\frac{1}{\max\limits_i k_{4,i}^2}\frac{t^2}{n}\right) & \text{if } n > \frac{t}{\max\limits_i k_{4,i}}\\ 2\exp\left(-\frac{1}{\max\limits_i k_{4,i}^2} t\right) & \text{if } n < \frac{t}{\max\limits_i k_{4,i}}, \end{cases}
\end{aligned} \tag{18}
$$

where the $k_{4,i}$'s are positive constants.

- Proof (via MGF):

$$\Pr\left\{\sum_{i=1}^{n} X_i \geq t\right\} = \Pr\left\{e^{\lambda\sum_{i=1}^{n} X_i} \geq e^{\lambda t}\right\} \leq \frac{E\left[e^{\lambda\sum_{i=1}^{n} X_i}\right]}{e^{\lambda t}}$$

$$= \frac{\prod_{i=1}^{n} E\left[e^{\lambda X_i}\right]}{e^{\lambda t}} \leq \frac{e^{\lambda^2\sum_{i=1}^{n} k_{4,i}^2}}{e^{\lambda t}} = e^{\lambda^2\sum_{i=1}^{n} k_{4,i}^2 - \lambda t} \tag{19}$$

We now need to maximize the right-hand side w.r.t. λ, however, λ is constrained to $|\lambda| \leq \min\limits_i \frac{1}{k_{4,i}} = \frac{1}{\max\limits_i k_{4,i}}$. Doing constrained optimization leads to the optimal choice of λ, given by

$$\lambda = \min\left\{\frac{t}{2\sum_{i=1}^{n} k_{4,i}^2}, \frac{1}{\max\limits_i k_{4,i}}\right\} \tag{20}$$

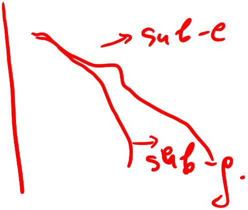
Substituting (20) into (19), we obtained the one sided Bernstein's Inequality as

$$\Pr\left\{\sum_{i=1}^{n} X_i \geq t\right\} \leq \exp\left(-\min\left\{\frac{t^2}{\sum_{i=1}^{n} k_{1,i}^2}, \frac{t}{\max\limits_i k_{1,i}}\right\}\right)$$
$$\leq \exp\left(-\min\left\{\frac{t^2}{n\max\limits_i k_{1,i}^2}, \frac{t}{\max\limits_i k_{1,i}}\right\}\right) \tag{21}$$

Following, the same procedure for deriving $\Pr\left\{-\sum_{i=1}^{n} X_i \geq t\right\}$, we again reach (21). Therefore, $\Pr\left\{\left|\sum_{i=1}^{n} X_i\right| \geq t\right\}$ is simply two times (21). Q.E.D.

Sub-Exponential RVs

- What is very important in the Bernstein's Inequality, given by (18), is that
  - For $n < \frac{t}{\max_i k_{4,i}}$, the tail of the sub-exponential sum goes towards zero exponentially with t
  - For $n > \frac{t}{\max_i k_{4,i}}$, the tail of the sub-exponential sum goes towards zero exponentially with t^2, which is identical to the rate of the decay of the sub-Gaussian sum.



The Thin Shell Phenomenon of Sub-Gaussian Vectors

- Let $\boldsymbol{X}$ be n-dimensional vector given by $\boldsymbol{X} = [X_1, X_2, ..., X_n]$.
- The Euclidean norm of a random vector $\boldsymbol{X}$ is given by

$$||\boldsymbol{X}||_2 = \sqrt{\boldsymbol{X}\boldsymbol{X}^T} = \sqrt{\sum_{i=1}^{n} X_i^2}, \tag{22}$$

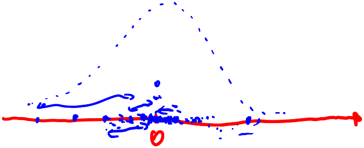
  where $(\cdot)^T$ denotes the transpose operation.
- Note that, for a vector $\boldsymbol{X}$, its norm, $||\boldsymbol{X}||_2$ represents the length of the vector $\boldsymbol{X}$, or equivalently its magnitude.
- The concentration inequalities that we have studied so far will allow us to study lengths of random vectors with independent components relatively easy.
- The Thin Shell Phenomenon of Sub-Gaussian Vectors will provide us with new insight on how we should visualize the lengths (i.e., norms) of high-dimensional sub-Gaussian vectors.

The Thin Shell Phenomenon of Sub-Gaussian Vectors

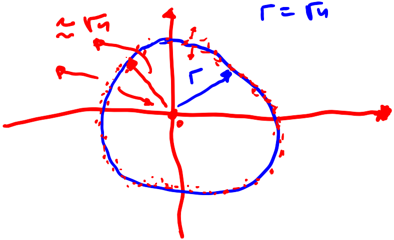
Lengths of Gaussian vectors:

- Let $\boldsymbol{X} = [X_1, X_2, ..., X_n]$ be a Gaussian vector comprised of elements X_i that are i.i.d. zero-mean unit-variance Gaussian RVs. Then the length/magnitude of this vector is given by (22).
- Now let's compare the lengths of $\boldsymbol{X}$ for the case when the dimension n is $n = 1$ to the case when n is very high.

a) When $n = 1$



b) When n is very high



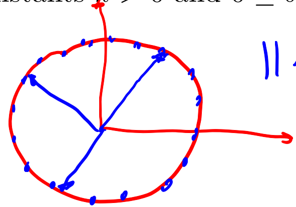
The Thin Shell Phenomenon of Sub-Gaussian Vectors

- Thm (Thin Shell): If the vector $\boldsymbol{X} = [X_1, X_2, ..., X_n]$ is comprised of elements that are i.i.d. sub-Gaussian RVs with $E[X_i^2] = 1$, then

$r = \sqrt{n}$

$$\Pr\left\{(1+\delta)\sqrt{n} \le ||\boldsymbol{X}||_2 \le (1-\delta)\sqrt{n}\right\} \le 2\exp\left(-n\frac{\delta(2-\delta)}{k}\min\left\{\frac{\delta(2-\delta)}{k}, 1\right\}\right) \tag{23}$$

for some constants $k > 0$ and $0 \le \delta \le 1$.



$||X||_2 = \sqrt{n} \pm \varepsilon$

$\varepsilon \to 0$ vary fast as $n \to \infty$

The Thin Shell Phenomenon of Sub-Gaussian Vectors

- Proof: Note that

$$\Pr\left\{(1+\delta)\sqrt{n} \le ||\boldsymbol{X}||_2 \le (1-\delta)\sqrt{n}\right\}$$
$$= \Pr\left\{||\boldsymbol{X}||_2 \ge (1+\delta)\sqrt{n}\right\} + \Pr\left\{||\boldsymbol{X}||_2 \le (1-\delta)\sqrt{n}\right\} \qquad (24)$$

since the events $||\boldsymbol{X}||_2 \ge (1+\delta)$ and $||\boldsymbol{X}||_2 \le (1-\delta)\sqrt{n}$ are disjoint.

Hence, to prove the Thm, we need to separately derive $\Pr\{||\boldsymbol{X}||_2 \ge (1+\delta)\sqrt{n}\}$ and $\Pr\{||\boldsymbol{X}||_2 \le (1-\delta)\sqrt{n}\}$ and then sum them. This is done in the following pages.

The Thin Shell Phenomenon of Sub-Gaussian Vectors

- Proof Continuation:

$$
\begin{aligned}
&\Pr\left\{||\boldsymbol{X}||_2 \geq (1+\delta)\sqrt{n}\right\} = \Pr\left\{||\boldsymbol{X}||_2^2 \geq (1+\delta)^2 n\right\} \\
&= \Pr\left\{||\boldsymbol{X}||_2^2 \geq n + 2\delta n + \delta^2 n\right\} = \Pr\left\{\sum_{i=1}^{n} X_i^2 - n \geq n\delta(2+\delta)\right\} \\
&= \Pr\left\{\sum_{i=1}^{n}(X_i^2 - 1) \geq n\delta(2+\delta)\right\} \overset{(a)}{=} \Pr\left\{\sum_{i=1}^{n} Y_i \geq n\delta(2+\delta)\right\} \\
&\overset{(b)}{\leq} \exp\left(-n\frac{\delta(2+\delta)}{k}\min\left\{\frac{\delta(2+\delta)}{k}, 1\right\}\right)
\end{aligned}
\tag{25}
$$

where (a) comes by making the substitution $Y_i = X_i^2 - 1$ and (b) follows by applying Bernstein's Inequality which is valid since Y_i is sub-exponential and $E[Y_i] = E[X_i^2] - 1 = 0$. Note that since Y_i's are i.i.d., the constants $k_{4,i}$ in Bernstein's Inequality satisfy $k_{4,i} = k$, $\forall i$.

The Thin Shell Phenomenon of Sub-Gaussian Vectors

Proof Continuation: Using the same technique, we can derive

$$\begin{aligned}
&\Pr\left\{||\boldsymbol{X}||_2 \leq (1-\delta)\sqrt{n}\right\} = \Pr\left\{||\boldsymbol{X}||_2^2 \leq (1-\delta)^2 n\right\} \\
&= \Pr\left\{||\boldsymbol{X}||_2^2 \leq n - 2\delta n + \delta^2 n\right\} = \Pr\left\{\sum_{i=1}^{n} X_i^2 - n \leq -n\delta(2-\delta)\right\} \\
&= \Pr\left\{\sum_{i=1}^{n} (X_i^2 - 1) \leq -n\delta(2-\delta)\right\} = \Pr\left\{\sum_{i=1}^{n} (1 - X_i^2) \geq n\delta(2-\delta)\right\} \\
&\overset{(a)}{=} \Pr\left\{\sum_{i=1}^{n} Y_i \geq n\delta(2-\delta)\right\} \overset{(b)}{\leq} \exp\left(-n\frac{\delta(2-\delta)}{k}\min\left\{\frac{\delta(2-\delta)}{k}, 1\right\}\right)
\end{aligned} \tag{26}$$

where (a) comes by making the substitution $Y_i = 1 - X_i^2$ and (b) follows by applying Bernstein's Inequality which is valid since Y_i is sub-exponential and $E[Y_i] = E[X_i^2] - 1 = 0$.

The Thin Shell Phenomenon of Sub-Gaussian Vectors

Proof Continuation: Now, since

$$\delta(2-\delta) \leq \delta(2+\delta) \tag{27}$$

for $0 \leq \delta \leq 1$, (25), can be upper bounded as

$$\begin{aligned} \Pr\left\{||\boldsymbol{X}||_2 \geq (1+\delta)\sqrt{n}\right\} &\leq \exp\left(-n\frac{\delta(2+\delta)}{k}\min\left\{\frac{\delta(2+\delta)}{k}, 1\right\}\right) \\ &\leq \exp\left(-n\frac{\delta(2-\delta)}{k}\min\left\{\frac{\delta(2-\delta)}{k}, 1\right\}\right) \end{aligned} \tag{28}$$

Summing (28) and (26), we obtain (23). Q.E.D.

The Thin Shell Phenomenon of Sub-Gaussian Vectors

- What the Thin-Shell Thm tells us is that the length of $\boldsymbol{X}$ is very close to $\sqrt{n}$ with very high probability.
- In other words, what the The Thin Shell Phenomenon tells us is that high-dimensional vectors, comprised of n sub-Gaussian RVs, live very close to the surface of an n-dimensional sphere with radius $\sqrt{n}$.
- Moreover, it tells us that is extremely unlikely for any high-dimensional vector, comprises of n sub-Gaussian RVs, to live above or bellow the surface of the sphere with radius $\sqrt{n}$. In fact, the probability that we will find such a vector living at a distance δ above or bellow the surface of an n-dimensional sphere is

$$2\exp\left(-n\frac{\delta(2-\delta)}{k}\min\left\{\frac{\delta(2-\delta)}{k},1\right\}\right)\to 0,\ \text{as } n\to\infty$$