EMALCA 2025: Lecture 1

Oscar-Hernan Madrid-Padilla
Department of Statistics, UCLA
oscar.madrid@stat.ucla.edu

June 29, 2025

Personal Story

- Crecí en pequeños pueblos de **Honduras**, con poblaciones que iban de 300 a 1,000 personas: Concepción del Norte, Santa Bárbara; aldea El Carrizal, Olancho.

- Mis padres no tuvieron educación universitaria y durante años vivimos sin necesidades básicas como electricidad o agua potable.

Personal Story

- Nuestra escuela tenía solo un maestro, que enseñaba varios grados al mismo tiempo.
- La mayoría de los niños con los que crecí ahora son inmigrantes indocumentados, están desempleados o, trágicamente, han sido asesinados.
- Mis padres, especialmente mi papá, estaban obsesionados con la educación y las matemáticas. Para mis hermanos y para mí, las matemáticas fueron un escape, una forma de soñar más allá de nuestras circunstancias:

Personal Story

- A pesar de no tener un título universitario, por nuestros logros en matemáticas, mi papá llegó a ser conocido en mi pueblo como "El Señor de las Matemáticas":

- Mis dos hermanos menores, **José** y **Carlos**, también se beneficiaron enormemente al participar en el programa de olimpiadas de matemáticas. Hoy en día, son profesores de matemáticas y estadística—**José en Virginia Tech y Carlos en la Universidad de Washington en St. Louis.**

Personal Story

- Gracias a las olimpiadas de matemáticas, obtuve una beca para estudiar la licenciatura en matemáticas en el **CIMAT (México)**, donde obtuve el promedio más alto de mi generación. Luego recibí una beca para realizar un doctorado en estadística en **The University of Texas at Austin.**

- De 2017 a 2019 fui Profesor Visitante Neyman en **University of California, Berkeley**.

- Desde 2019 soy profesor en **University of California, Los Angeles** (La universidad pública número 1 de USA, consistentemente clasificada entre las mejores universidades a nivel mundial) y estoy a punto de postularme para la definitividad.

Personal Story

- Hago investigación en aprendizaje automático y estadística, y he publicado alrededor de 40 artículos en las principales revistas y conferencias del área. He coescrito differentes artículos con Daniela Witten y Ryan Tibshirani, dos ganadores del Premio COPSS Presidents' Award, uno de los reconocimientos más prestigiosos en el campo de la estadística.

- Algunas personas geniales que conocí:

(a) Con mi hermano José y mi colega de UCLA el medallista Fields Terence Tao.

(b) Con Kip Thorne, Premio Nobel de Física 2017.

References

- Wainwright, Martin J. "High-dimensional statistics." Camb. Ser. Stat. Probab. Math 48 (1945). (HDS).
- Vershynin, Roman. High-dimensional probability: An introduction with applications in data science. Vol. 47. Cambridge university press, 2018. (HDP). **https://www.math.uci.edu/ rvershyn/papers/HDP-book/HDP-2.pdf.**

- Classical case:
  - $n >> d$.

- Classical case:
    - $n >> d$.
    - Asymptotic assumption: d is fixed and $n \to \infty$.

- Classical case:
  - $n >> d$.
  - Asymptotic assumption: d is fixed and $n \to \infty$.
  - Basic tools: LLN and CLT.

- Classical case:
  - $n >> d$.
  - Asymptotic assumption: d is fixed and $n \to \infty$.
  - Basic tools: LLN and CLT.
-
- High-dimensional setting:

- Classical case:
  - $n >> d$.
  - Asymptotic assumption: d is fixed and $n \to \infty$.
  - Basic tools: LLN and CLT.
-
- High-dimensional setting:
  - $n \asymp d$, e.g. $d/n \to \gamma$ or

- Classical case:
  - $n >> d$.
  - Asymptotic assumption: d is fixed and $n \to \infty$.
  - Basic tools: LLN and CLT.
-
- High-dimensional setting:
  - $n \asymp d$, e.g. $d/n \to \gamma$ or
  - $d >> n$, e.g. $d = e^n$.

- Classical case:
  - $n >> d$.
  - Asymptotic assumption: d is fixed and $n \to \infty$.
  - Basic tools: LLN and CLT.
-
- High-dimensional setting:
  - $n \asymp d$, e.g. $d/n \to \gamma$ or
  - $d >> n$, e.g. $d = e^n$.
  - E.g. 10^4 genes with only 50 samples.

- Classical case:
  - $n >> d$.
  - Asymptotic assumption: d is fixed and $n \to \infty$.
  - Basic tools: LLN and CLT.
-
- High-dimensional setting:
  - $n \asymp d$, e.g. $d/n \to \gamma$ or
  - $d >> n$, e.g. $d = e^n$.
  - E.g. 10^4 genes with only 50 samples.
- Classical methods fail. E.g.,

- Classical case:
  - $n >> d$.
  - Asymptotic assumption: d is fixed and $n \to \infty$.
  - Basic tools: LLN and CLT.
-
- High-dimensional setting:
  - $n \asymp d$, e.g. $d/n \to \gamma$ or
  - $d >> n$, e.g. $d = e^n$.
  - E.g. 10^4 genes with only 50 samples.
- Classical methods fail. E.g.,
- Linear regression $y = X\beta + \varepsilon$, where $\varepsilon \sim N(0, I_n)$.

$$\hat{\beta}_{ols} = \arg\min_{\beta \in \mathbb{R}^d} \|y - X\beta\|_2^2$$

- Classical case:
  - $n >> d$.
  - Asymptotic assumption: d is fixed and $n \to \infty$.
  - Basic tools: LLN and CLT.
-
- High-dimensional setting:
  - $n \asymp d$, e.g. $d/n \to \gamma$ or
  - $d >> n$, e.g. $d = e^n$.
  - E.g. 10^4 genes with only 50 samples.
- Classical methods fail. E.g.,
- Linear regression $y = X\beta + \varepsilon$, where $\varepsilon \sim N(0, I_n)$.

$$\hat{\beta}_{ols} = \arg\min_{\beta \in \mathbb{R}^d} \|y - X\beta\|_2^2$$

This is not even well-defined!

Concentration inequalities

- Main tools in dealing with high-dimensional randomness.

Concentration inequalities

- Main tools in dealing with high-dimensional randomness.
- Non-asymptotic versions of the CLT.

Concentration inequalities

- Main tools in dealing with high-dimensional randomness.
- Non-asymptotic versions of the CLT.
- General form: $\mathbb{P}(|X - \mathbb{E}(X)| \geq A) < B =$ something small.

Concentration inequalities

- Main tools in dealing with high-dimensional randomness.
- Non-asymptotic versions of the CLT.
- General form: $\mathbb{P}(|X - \mathbb{E}(X)| \geq A) < B =$ something small.
- Classical examples: Markov and Chebyshev inequalities:

Concentration inequalities

- Main tools in dealing with high-dimensional randomness.
- Non-asymptotic versions of the CLT.
- General form: $\mathbb{P}(|X - \mathbb{E}(X)| \geq A) < B =$ something small.
- Classical examples: Markov and Chebyshev inequalities:
- Markov: Assume that $X \geq 0$, then

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Concentration inequalities

- Main tools in dealing with high-dimensional randomness.
- Non-asymptotic versions of the CLT.
- General form: $\mathbb{P}(|X - \mathbb{E}(X)| \geq A) < B =$ something small.
- Classical examples: Markov and Chebyshev inequalities:
- Markov: Assume that $X \geq 0$, then

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

- Chebyshev: Assume $\mathbb{E}(X^2) < \infty$, and let $\mu = \mathbb{E}(X)$. Then

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \ \leq \ \frac{\mathrm{Var}(X)}{t^2},$$

Concentration inequalities

- Main tools in dealing with high-dimensional randomness.
- Non-asymptotic versions of the CLT.
- General form: $\mathbb{P}(|X - \mathbb{E}(X)| \geq A) < B =$ something small.
- Classical examples: Markov and Chebyshev inequalities:
- Markov: Assume that $X \geq 0$, then

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

- Chebyshev: Assume $\mathbb{E}(X^2) < \infty$, and let $\mu = \mathbb{E}(X)$. Then

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \; \leq \; \frac{\mathrm{Var}(X)}{t^2},$$

- Stronger assumption: $\mathbb{E}(|X|^k) < \infty$, and let $\mu = \mathbb{E}(X)$. Then

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \; \leq \; \frac{\mathbb{E}\left(|X - \mu|^k\right)}{t^k}.$$

Example 1

- Let $X_1, \ldots, X_n \sim \text{Ber}(1/2)$ and $S_n = \sum_{i=1}^n X_i$. Then, by CLT,

$$Z_n := \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - n/2}{\sqrt{n/4}} \xrightarrow{\text{d}} N(0,1).$$

Example 1

- Let $X_1, \ldots, X_n \sim \text{Ber}(1/2)$ and $S_n = \sum_{i=1}^n X_i$. Then, by CLT,

$$Z_n := \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - n/2}{\sqrt{n/4}} \xrightarrow{\text{d}} N(0, 1).$$

- Letting $g \sim N(0, 1)$,

$$\mathbb{P}\left(S_n \geq \frac{n}{2} + \sqrt{\frac{n}{4}}\,t\right) = \mathbb{P}\left(\frac{S_n - n/2}{\sqrt{n/4}} \geq t\right)$$

Example 1

- Let $X_1, \dots, X_n \sim \mathrm{Ber}(1/2)$ and $S_n = \sum_{i=1}^n X_i$. Then, by CLT,

$$Z_n := \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - n/2}{\sqrt{n/4}} \xrightarrow{\mathrm{d}} N(0,1).$$

- Letting $g \sim N(0,1)$,

$$\mathbb{P}\left(S_n \geq \frac{n}{2} + \sqrt{\frac{n}{4}}\,t\right) = \mathbb{P}\left(\frac{S_n - n/2}{\sqrt{n/4}} \geq t\right) \approx \mathbb{P}(g \geq t) \leq \frac{e^{-t^2/2}}{2}.$$

Example 1

- Let $X_1, \ldots, X_n \sim \text{Ber}(1/2)$ and $S_n = \sum_{i=1}^n X_i$. Then, by CLT,

$$Z_n := \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - n/2}{\sqrt{n/4}} \xrightarrow{\text{d}} N(0,1).$$

- Letting $g \sim N(0,1)$,

$$\mathbb{P}\left(S_n \geq \frac{n}{2} + \sqrt{\frac{n}{4}}t\right) = \mathbb{P}\left(\frac{S_n - n/2}{\sqrt{n/4}} \geq t\right) \approx \mathbb{P}(g \geq t) \leq \frac{e^{-t^2/2}}{2}.$$

- Letting $t = \alpha\sqrt{n}$,

$$\mathbb{P}(S_n \geq \frac{n}{2}(1+\alpha)) \leq \frac{1}{2}e^{-n\alpha^2/2}.$$

Example 1

- Let $X_1, \ldots, X_n \sim \text{Ber}(1/2)$ and $S_n = \sum_{i=1}^n X_i$. Then, by CLT,

$$Z_n := \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - n/2}{\sqrt{n/4}} \xrightarrow{\text{d}} N(0,1).$$

- Letting $g \sim N(0,1)$,

$$\mathbb{P}\left(S_n \geq \frac{n}{2} + \sqrt{\frac{n}{4}}t\right) = \mathbb{P}\left(\frac{S_n - n/2}{\sqrt{n/4}} \geq t\right) \approx \mathbb{P}(g \geq t) \leq \frac{e^{-t^2/2}}{2}.$$

- Letting $t = \alpha\sqrt{n}$,

$$\mathbb{P}(S_n \geq \frac{n}{2}(1+\alpha)) \leq \frac{1}{2}e^{-n\alpha^2/2}.$$

- Problem: Approximation is not tight in general.

- **Theorem 1. (Berry–Esseen CLT)** Under the assumption of CLT, with $\rho = \mathbb{E}\left(|X_1 - \mu|^3\right)/\sigma^3$,

$$\sup_{t\in\mathbb{R}} |\mathbb{P}(Z_n \geq t) - \mathbb{P}(g \geq t)| \leq \frac{C\rho}{n^{1/2}},$$

for a positive constant C, where $\sigma^2 = \mathrm{Var}(X_1)$.

- **Theorem 1. (Berry–Esseen CLT)** Under the assumption of CLT, with $\rho = \mathbb{E}\left(|X_1 - \mu|^3\right)/\sigma^3$,

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(Z_n \geq t) - \mathbb{P}(g \geq t)| \leq \frac{C\rho}{n^{1/2}},$$

  for a positive constant C, where $\sigma^2 = \mathrm{Var}(X_1)$.
- Conclusion, the approximation error is $O(n^{-1/2})$ which is a lot larger than the exponential bound $O(e^{-n\alpha^2/2})$ that we want to establish.

- **Theorem 1. (Berry–Esseen CLT)** Under the assumption of CLT, with $\rho = \mathbb{E}\left(|X_1 - \mu|^3\right)/\sigma^3$,

$$\sup_{t\in\mathbb{R}} |\mathbb{P}(Z_n \geq t) - \mathbb{P}(g \geq t)| \leq \frac{C\rho}{n^{1/2}},$$

  for a positive constant C, where $\sigma^2 = \mathrm{Var}(X_1)$.
- Conclusion, the approximation error is $O(n^{-1/2})$ which is a lot larger than the exponential bound $O(e^{-n\alpha^2/2})$ that we want to establish.
- Solution: Directly obtain the concentration inequalities.

- **Theorem 1. (Berry–Esseen CLT)** Under the assumption of CLT, with $\rho = \mathbb{E}\left(|X_1 - \mu|^3\right)/\sigma^3$,

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(Z_n \geq t\right) - \mathbb{P}(g \geq t) \right| \leq \frac{C\rho}{n^{1/2}},$$

  for a positive constant C, where $\sigma^2 = \mathrm{Var}(X_1)$.
- Conclusion, the approximation error is $O(n^{-1/2})$ which is a lot larger than the exponential bound $O(e^{-n\alpha^2/2})$ that we want to establish.
- Solution: Directly obtain the concentration inequalities.
- Often using Chernoff bounding technique: for any $\lambda > 0$

$$\mathbb{P}\left(Z_n \geq t\right) = \mathbb{P}\left(e^{\lambda Z_n} \geq e^{\lambda t}\right)$$

- **Theorem 1. (Berry–Esseen CLT)** Under the assumption of CLT, with $\rho = \mathbb{E}\left(|X_1 - \mu|^3\right)/\sigma^3$,

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(Z_n \geq t) - \mathbb{P}(g \geq t)| \leq \frac{C\rho}{n^{1/2}},$$

  for a positive constant C, where $\sigma^2 = \mathrm{Var}(X_1)$.
- Conclusion, the approximation error is $O(n^{-1/2})$ which is a lot larger than the exponential bound $O(e^{-n\alpha^2/2})$ that we want to establish.
- Solution: Directly obtain the concentration inequalities.
- Often using Chernoff bounding technique: for any $\lambda > 0$

$$\mathbb{P}(Z_n \geq t) = \mathbb{P}\left(e^{\lambda Z_n} \geq e^{\lambda t}\right) \leq \frac{\mathbb{E}\left(e^{\lambda Z_n}\right)}{e^{\lambda t}}, \quad t \in \mathbb{R}.$$

- **Theorem 1. (Berry–Esseen CLT)** Under the assumption of CLT, with $\rho = \mathbb{E}\left(|X_1 - \mu|^3\right)/\sigma^3$,

$$\sup_{t \in \mathbb{R}} \left|\mathbb{P}\left(Z_n \geq t\right) - \mathbb{P}(g \geq t)\right| \leq \frac{C\rho}{n^{1/2}},$$

  for a positive constant C, where $\sigma^2 = \mathrm{Var}(X_1)$.
- Conclusion, the approximation error is $O(n^{-1/2})$ which is a lot larger than the exponential bound $O(e^{-n\alpha^2/2})$ that we want to establish.
- Solution: Directly obtain the concentration inequalities.
- Often using Chernoff bounding technique: for any $\lambda > 0$

$$\mathbb{P}\left(Z_n \geq t\right) = \mathbb{P}\left(e^{\lambda Z_n} \geq e^{\lambda t}\right) \leq \frac{\mathbb{E}\left(e^{\lambda Z_n}\right)}{e^{\lambda t}}, \quad t \in \mathbb{R}.$$

- Leads to the study of the moment generating function (MGF) of random variables.

Sub-Gaussian concentration

- **Definition 1.** A zero-mean random variable X is sub-Gaussian if for some $\sigma > 0$

$$\mathbb{E}\left(e^{\lambda X}\right) \leq e^{\sigma^2 \lambda^2 / 2} \quad \forall \lambda \in \mathbb{R}. \tag{1}$$

  A general random variable is sub-Gaussian if $X - \mathbb{E}(X)$ is sub-Gaussian.

Sub-Gaussian concentration

- **Definition 1.** A zero-mean random variable X is sub-Gaussian if for some $\sigma > 0$

$$\mathbb{E}\left(e^{\lambda X}\right) \leq e^{\sigma^2 \lambda^2 / 2} \quad \forall \lambda \in \mathbb{R}. \tag{1}$$

  A general random variable is sub-Gaussian if $X - \mathbb{E}(X)$ is sub-Gaussian.
- $X \sim N(0, \sigma^2)$ satisfies Definition 1 with equality.

Sub-Gaussian concentration

- **Definition 1.** A zero-mean random variable X is sub-Gaussian if for some $\sigma > 0$

$$\mathbb{E}\left(e^{\lambda X}\right) \leq e^{\sigma^2\lambda^2/2} \quad \forall \lambda \in \mathbb{R}. \tag{1}$$

  A general random variable is sub-Gaussian if $X - \mathbb{E}(X)$ is sub-Gaussian.
- $X \sim N(0, \sigma^2)$ satisfies Definition 1 with equality.
- A Rademacher variable, also called symmetric Bernoulli $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$ is sub-Gaussian,

$$\mathbb{E}(e^{\lambda X}) = \cosh(\lambda) \leq e^{\lambda^2/2}.$$

Sub-Gaussian concentration

- **Definition 1.** A zero-mean random variable X is sub-Gaussian if for some $\sigma > 0$

$$\mathbb{E}\left(e^{\lambda X}\right) \le e^{\sigma^2\lambda^2/2} \quad \forall \lambda \in \mathbb{R}. \tag{1}$$

  A general random variable is sub-Gaussian if $X - \mathbb{E}(X)$ is sub-Gaussian.
- $X \sim N(0, \sigma^2)$ satisfies Definition 1 with equality.
- A Rademacher variable, also called symmetric Bernoulli $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$ is sub-Gaussian,

$$\mathbb{E}(e^{\lambda X}) = \cosh(\lambda) \le e^{\lambda^2/2}.$$

- Any bounded RV is sub-Gaussian: $X \in [a, b]$ a.s., then (1) holds with $\sigma = \frac{b-a}{2}$.

- **Proposition 1.** Assume that X is zero-mean sub-Gaussian satisfying (1). Then,

$$\mathbb{P}(X \geq t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right), \quad \forall t \geq 0.$$

  Same bound holds with X replaced with $-X$.

- **Proposition 1.** Assume that X is zero-mean sub-Gaussian satisfying (1). Then,

  $$\mathbb{P}(X \geq t) \;\leq\; \exp\left(-\frac{t^2}{2\sigma^2}\right), \quad \forall t \geq 0.$$

  Same bound holds with X replaced with $-X$.
  Proof. Chernoff bound

  $$\mathbb{P}(X \geq t) \;\leq\; \inf_{\lambda > 0}\left[e^{-\lambda t}\mathbb{E}\left(e^{\lambda X}\right)\right]$$

- **Proposition 1.** Assume that X is zero-mean sub-Gaussian satisfying (1). Then,

$$\mathbb{P}(X \geq t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right), \quad \forall t \geq 0.$$

Same bound holds with X replaced with $-X$.
Proof. Chernoff bound

$$\mathbb{P}(X \geq t) \leq \inf_{\lambda > 0} \left[e^{-\lambda t}\mathbb{E}\left(e^{\lambda X}\right)\right] \leq \inf_{\lambda > 0} \exp\left(-\lambda t + \frac{\lambda^2\sigma^2}{2}\right).$$

- **Proposition 1.** Assume that X is zero-mean sub-Gaussian satisfying (1). Then,

  $$\mathbb{P}(X \geq t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right), \quad \forall t \geq 0.$$

  Same bound holds with X replaced with $-X$.
  Proof. Chernoff bound

  $$\mathbb{P}(X \geq t) \leq \inf_{\lambda > 0} \left[e^{-\lambda t}\mathbb{E}\left(e^{\lambda X}\right)\right] \leq \inf_{\lambda > 0} \exp\left(-\lambda t + \frac{\lambda^2\sigma^2}{2}\right).$$

- Union bound gives two-sided bound:
  $\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right)$

- **Proposition 1.** Assume that X is zero-mean sub-Gaussian satisfying (1). Then,

$$\mathbb{P}(X \geq t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right), \quad \forall t \geq 0.$$

  Same bound holds with X replaced with $-X$.
  Proof. Chernoff bound

$$\mathbb{P}(X \geq t) \leq \inf_{\lambda>0}\left[e^{-\lambda t}\mathbb{E}\left(e^{\lambda X}\right)\right] \leq \inf_{\lambda>0}\exp\left(-\lambda t + \frac{\lambda^2\sigma^2}{2}\right).$$

- Union bound gives two-sided bound:
  $\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right)$
- What if $\mathbb{E}(X) \neq 0$?

- **Proposition 1.** Assume that X is zero-mean sub-Gaussian satisfying (1). Then,

$$\mathbb{P}(X \geq t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right), \quad \forall t \geq 0.$$

  Same bound holds with X replaced with $-X$.
  Proof. Chernoff bound

$$\mathbb{P}(X \geq t) \leq \inf_{\lambda > 0}\left[e^{-\lambda t}\mathbb{E}\left(e^{\lambda X}\right)\right] \leq \inf_{\lambda > 0} \exp\left(-\lambda t + \frac{\lambda^2\sigma^2}{2}\right).$$

- Union bound gives two-sided bound: $\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right)$
- What if $\mathbb{E}(X) \neq 0$?
- Apply inequality to $X - \mathbb{E}(X)$ to obtain

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right).$$

- **Proposition 2.** Assume that $\{X_i\}$ are independent zero-mean sub-Gaussian with parameters $\{\sigma_i\}$. Then, $S_n = \sum_{i=1}^n X_i$ is sub-Gaussian with parameter $\sigma := \sqrt{\sum_{i=1}^n \sigma_i^2}$.

- **Proposition 2.** Assume that $\{X_i\}$ are independent zero-mean sub-Gaussian with parameters $\{\sigma_i\}$. Then, $S_n = \sum_{i=1}^n X_i$ is sub-Gaussian with parameter $\sigma := \sqrt{\sum_{i=1}^n \sigma_i^2}$.
- **Proof.**

$$\mathbb{E}\left(e^{\lambda S_n}\right) = \mathbb{E}\left(e^{\lambda \sum_{i=1}^n X_i}\right)$$

- **Proposition 2.** Assume that $\{X_i\}$ are independent zero-mean sub-Gaussian with parameters $\{\sigma_i\}$. Then, $S_n = \sum_{i=1}^n X_i$ is sub-Gaussian with parameter $\sigma := \sqrt{\sum_{i=1}^n \sigma_i^2}$.
- **Proof.**

$$\begin{aligned}
\mathbb{E}\left(e^{\lambda S_n}\right) &= \mathbb{E}\left(e^{\lambda \sum_{i=1}^n X_i}\right) \\
&= \mathbb{E}\left(\prod_{i=1}^n e^{\lambda X_i}\right)
\end{aligned}$$

- **Proposition 2.** Assume that $\{X_i\}$ are independent zero-mean sub-Gaussian with parameters $\{\sigma_i\}$. Then, $S_n = \sum_{i=1}^n X_i$ is sub-Gaussian with parameter $\sigma := \sqrt{\sum_{i=1}^n \sigma_i^2}$.
- **Proof.**

$$\begin{aligned} \mathbb{E}\left(e^{\lambda S_n}\right) &= \mathbb{E}\left(e^{\lambda \sum_{i=1}^n X_i}\right) \\ &= \mathbb{E}\left(\prod_{i=1}^n e^{\lambda X_i}\right) \\ &= \prod_{i=1}^n \mathbb{E}\left(e^{\lambda X_i}\right) \end{aligned}$$

- **Proposition 2.** Assume that $\{X_i\}$ are independent zero-mean sub-Gaussian with parameters $\{\sigma_i\}$. Then, $S_n = \sum_{i=1}^n X_i$ is sub-Gaussian with parameter $\sigma := \sqrt{\sum_{i=1}^n \sigma_i^2}$.
- **Proof.**

$$\begin{aligned}
\mathbb{E}\left(e^{\lambda S_n}\right) &= \mathbb{E}\left(e^{\lambda \sum_{i=1}^n X_i}\right) \\
&= \mathbb{E}\left(\prod_{i=1}^n e^{\lambda X_i}\right) \\
&= \prod_{i=1}^n \mathbb{E}\left(e^{\lambda X_i}\right) \\
&\leq \prod_{i=1}^n e^{\lambda^2 \sigma_i^2/2}
\end{aligned}$$

- **Proposition 2.** Assume that $\{X_i\}$ are independent zero-mean sub-Gaussian with parameters $\{\sigma_i\}$. Then, $S_n = \sum_{i=1}^n X_i$ is sub-Gaussian with parameter $\sigma := \sqrt{\sum_{i=1}^n \sigma_i^2}$.
- **Proof.**

$$\begin{aligned}
\mathbb{E}\left(e^{\lambda S_n}\right) &= \mathbb{E}\left(e^{\lambda \sum_{i=1}^n X_i}\right) \\
&= \mathbb{E}\left(\prod_{i=1}^n e^{\lambda X_i}\right) \\
&= \prod_{i=1}^n \mathbb{E}\left(e^{\lambda X_i}\right) \\
&\leq \prod_{i=1}^n e^{\lambda^2 \sigma_i^2/2} \\
&= e^{\lambda^2 \left(\sum_{i=1}^n \sigma_i^2\right)/2}
\end{aligned}$$

- **Theorem 2 (Hoeffding).** Assume that $\{X_i\}$ are independent zero-mean sub-Gaussian with parameters $\{\sigma_i\}$ and let $\sigma := \sqrt{\sum_{i=1}^n \sigma_i^2}$. Then

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right), \quad t \geq 0.$$

  Same bound holds with X_i replaced with $-X_i$.

- **Theorem 2 (Hoeffding).** Assume that $\{X_i\}$ are independent zero-mean sub-Gaussian with parameters $\{\sigma_i\}$ and let $\sigma := \sqrt{\sum_{i=1}^n \sigma_i^2}$. Then

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right), \quad t \geq 0.$$

  Same bound holds with X_i replaced with $-X_i$.
- Alternative form: Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $\bar{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \sigma_i^2$. Then

$$\mathbb{P}\left(\bar{X}_n \geq t\right) \leq \exp\left(-\frac{nt^2}{2\bar{\sigma}^2}\right), \quad t \geq 0.$$

- **Theorem 2 (Hoeffding).** Assume that $\{X_i\}$ are independent zero-mean sub-Gaussian with parameters $\{\sigma_i\}$ and let $\sigma := \sqrt{\sum_{i=1}^n \sigma_i^2}$. Then

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right), \quad t \geq 0.$$

  Same bound holds with X_i replaced with $-X_i$.
- Alternative form: Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $\bar{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \sigma_i^2$. Then

$$\mathbb{P}\left(\bar{X}_n \geq t\right) \leq \exp\left(-\frac{nt^2}{2\bar{\sigma}^2}\right), \quad t \geq 0.$$

- Example: $X_i \sim \text{Rad}$ so that $\sigma_i = 1$ and $\bar{\sigma} = 1$, then

$$\mathbb{P}\left(\bar{X}_n \geq \frac{t}{\sqrt{n}}\right) \leq \exp\left(-\frac{t^2}{2}\right)$$

Equivalent characterizations of sub-Gaussianity

For a random variable X, the following are equivalent: (HDP, Prop. 2.6). Below $K_1, \ldots, K_4$ are positive constants.

1. The tails of X satisfy

$$\mathbb{P}\left(|X| \geq t\right) \leq 2\exp\left(-\frac{t^2}{K_1^2}\right), \quad \forall t \geq 0.$$

Equivalent characterizations of sub-Gaussianity

For a random variable X, the following are equivalent: (HDP, Prop. 2.6). Below $K_1, \ldots, K_4$ are positive constants.

1. The tails of X satisfy
$$\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{K_1^2}\right), \quad \forall t \geq 0.$$
2. The moments of X satisfy
$$\|X\|_p = \left(\mathbb{E}(|X|^p)\right)^{1/p} \leq K_2\sqrt{p}, \quad \forall p \geq 1.$$

Equivalent characterizations of sub-Gaussianity

For a random variable X, the following are equivalent: (HDP, Prop. 2.6). Below $K_1, \ldots, K_4$ are positive constants.

1. The tails of X satisfy

$$\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{K_1^2}\right), \quad \forall t \geq 0.$$

2. The moments of X satisfy

$$\|X\|_p = \left(\mathbb{E}(|X|^p)\right)^{1/p} \leq K_2\sqrt{p}, \quad \forall p \geq 1.$$

3. The MGF of X^2 satisfies

$$\mathbb{E}\left(e^{\lambda^2 X^2}\right) \leq \exp\left(K_3^2\lambda^2\right), \quad \forall|\lambda| \leq \frac{1}{K_3}.$$

Equivalent characterizations of sub-Gaussianity

For a random variable X, the following are equivalent: (HDP, Prop. 2.6). Below $K_1, \ldots, K_4$ are positive constants.

1. The tails of X satisfy
$$\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{K_1^2}\right), \quad \forall t \geq 0.$$
2. The moments of X satisfy
$$\|X\|_p = \left(\mathbb{E}(|X|^p)\right)^{1/p} \leq K_2\sqrt{p}, \quad \forall p \geq 1.$$
3. The MGF of X^2 satisfies
$$\mathbb{E}\left(e^{\lambda^2 X^2}\right) \leq \exp\left(K_3^2\lambda^2\right), \quad \forall |\lambda| \leq \frac{1}{K_3}.$$
4. The MGF of X^2 is bounded at some point,
$$\mathbb{E}\left(\exp\left(X^2/K_4\right)\right) \leq 2.$$

- Assuming $\mathbb{E}(X) = 0$, the above are equivalent to:

$$\mathbb{E}\left(\exp\left(\lambda X\right)\right) \quad \leq \quad \exp\left(\lambda^2 K_5^2\right), \quad \forall \lambda \in \mathbb{R}$$

Proof $\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{K_1^2}\right)$ **implies** $\|X\|_p = (\mathbb{E}(|X|^p))^{1/p} \leq K_2\sqrt{p}$.

- Rescalling X to X/K_1 we can assume that $K_1 = 1$. Then

$$\mathbb{E}(|X|^p) = \int_0^\infty \mathbb{P}(|X|^p \geq u)\, du$$

Proof $\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{K_1^2}\right)$ **implies** $\|X\|_p = (\mathbb{E}(|X|^p))^{1/p} \leq K_2\sqrt{p}$.

- Rescalling X to X/K_1 we can assume that $K_1 = 1$. Then

$$\begin{aligned} \mathbb{E}(|X|^p) &= \int_0^\infty \mathbb{P}(|X|^p \geq u)\, du \\ &= \int_0^\infty \mathbb{P}(|X| \geq t) p t^{p-1} dt \qquad (u = t^p,\ du = pt^{p-1} dt) \end{aligned}$$

Proof $\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{K_1^2}\right)$ **implies** $\|X\|_p = \left(\mathbb{E}(|X|^p)\right)^{1/p} \leq K_2\sqrt{p}$**.**

- Rescalling X to X/K_1 we can assume that $K_1 = 1$. Then

$$
\begin{aligned}
\mathbb{E}(|X|^p) &= \int_0^\infty \mathbb{P}(|X|^p \geq u)\, du \\
&= \int_0^\infty \mathbb{P}(|X| \geq t) p t^{p-1} dt \qquad (u = t^p,\ du = pt^{p-1} dt) \\
&\leq \int_0^\infty 2e^{-t^2} p t^{p-1} dt \qquad \text{by Property 1}
\end{aligned}
$$

Proof $\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{K_1^2}\right)$ **implies** $\|X\|_p = \left(\mathbb{E}(|X|^p)\right)^{1/p} \leq K_2\sqrt{p}$.

- Rescalling X to X/K_1 we can assume that $K_1 = 1$. Then

$$\begin{array}{rcl}
\mathbb{E}(|X|^p) & = & \int_0^\infty \mathbb{P}(|X|^p \geq u)\, du \\
& = & \int_0^\infty \mathbb{P}(|X| \geq t) p t^{p-1} dt \quad (u = t^p,\ du = pt^{p-1}dt) \\
& \leq & \int_0^\infty 2e^{-t^2} p t^{p-1} dt \quad \text{by Property 1} \\
& = & p\int_0^\infty e^{-s} s^{p/2-1} ds \quad \text{set } s = t^2,\ \ ds = 2tdt
\end{array}$$

Proof $\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{K_1^2}\right)$ **implies** $\|X\|_p = (\mathbb{E}(|X|^p))^{1/p} \leq K_2\sqrt{p}$.

- Rescalling X to X/K_1 we can assume that $K_1 = 1$. Then

$$
\begin{aligned}
\mathbb{E}(|X|^p) &= \int_0^\infty \mathbb{P}(|X|^p \geq u)\, du \\
&= \int_0^\infty \mathbb{P}(|X| \geq t) p t^{p-1} dt \qquad (u = t^p,\ du = p t^{p-1} dt) \\
&\leq \int_0^\infty 2e^{-t^2} p t^{p-1} dt \qquad \text{by Property 1} \\
&= p\int_0^\infty e^{-s} s^{p/2-1} ds \qquad \text{set } s = t^2,\ \ ds = 2t dt \\
&= p\Gamma(p/2)
\end{aligned}
$$

Proof $\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{K_1^2}\right)$ **implies** $\|X\|_p = \left(\mathbb{E}(|X|^p)\right)^{1/p} \leq K_2\sqrt{p}$.

- Rescalling X to X/K_1 we can assume that $K_1 = 1$. Then

$$
\begin{aligned}
\mathbb{E}(|X|^p) &= \int_0^\infty \mathbb{P}(|X|^p \geq u)\, du \\
&= \int_0^\infty \mathbb{P}(|X| \geq t) p t^{p-1} dt \quad (u = t^p,\ du = p t^{p-1} dt) \\
&\leq \int_0^\infty 2e^{-t^2} p t^{p-1} dt \quad \text{by Property 1} \\
&= p \int_0^\infty e^{-s} s^{p/2-1} ds \quad \text{set } s = t^2,\ \ ds = 2t dt \\
&= p\Gamma(p/2) \\
&\leq 3p(p/2)^{p/2} \quad \text{since } \Gamma(x) \leq 3x^x \ \text{ for all } x \geq 1/2
\end{aligned}
$$

Proof $\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{K_1^2}\right)$ **implies** $\|X\|_p = \left(\mathbb{E}(|X|^p)\right)^{1/p} \leq K_2\sqrt{p}$.

- Rescalling X to X/K_1 we can assume that $K_1 = 1$. Then

$$
\begin{array}{rcll}
\mathbb{E}(|X|^p) & = & \int_0^\infty \mathbb{P}(|X|^p \geq u)\, du & \\
& = & \int_0^\infty \mathbb{P}(|X| \geq t) p t^{p-1} dt & (u = t^p,\ du = pt^{p-1}dt) \\
& \leq & \int_0^\infty 2e^{-t^2} p t^{p-1} dt & \text{by Property 1} \\
& = & p\int_0^\infty e^{-s} s^{p/2-1} ds & \text{set } s = t^2,\ \ ds = 2tdt \\
& = & p\Gamma(p/2) & \\
& \leq & 3p(p/2)^{p/2} & \text{since } \Gamma(x) \leq 3x^x \text{ for all } x \geq 1/2
\end{array}
$$

- Taking the pth root root yields Property 2.

Proof $\left(\mathbb{E}\left(|X|^{p}\right)\right)^{1/p} \leq K_2\sqrt{p}$ **for all** $p \geq 1$ **implies** $\mathbb{E}\left(e^{\lambda^2 X^2}\right) \leq \exp\left(K_3^2\lambda^2\right), \quad \forall|\lambda| \leq \frac{1}{K_3}$**.**

Proof $(\mathbb{E}(|X|^p))^{1/p} \leq K_2\sqrt{p}$ **for all** $p \geq 1$ **implies** $\mathbb{E}\left(e^{\lambda^2 X^2}\right) \leq \exp\left(K_3^2\lambda^2\right), \quad \forall|\lambda| \leq \frac{1}{K_3}$**.**

- Assume property ii holds. As before, by homogeneity we may assume that $K_2 = 1$. Recalling the Taylor series expansion of the exponential function, we obtain

$$.$$

Proof $(\mathbb{E}(|X|^p))^{1/p} \le K_2\sqrt{p}$ **for all** $p \ge 1$ **implies** $\mathbb{E}\left(e^{\lambda^2 X^2}\right) \le \exp\left(K_3^2\lambda^2\right), \quad \forall|\lambda| \le \frac{1}{K_3}$**.**

- Assume property ii holds. As before, by homogeneity we may assume that $K_2 = 1$. Recalling the Taylor series expansion of the exponential function, we obtain

$$\mathbb{E}\left(e^{\lambda^2 X^2}\right) = \mathbb{E}\left(1 + \sum_{p=1}^{\infty} \frac{(\lambda^2 X^2)^p}{p!}\right)$$

.

Proof $\left(\mathbb{E}\left(|X|^p\right)\right)^{1/p} \le K_2\sqrt{p}$ for all $p \geq 1$ implies $\mathbb{E}\left(e^{\lambda^2X^2}\right) \le \exp\left(K_3^2\lambda^2\right), \quad \forall|\lambda| \le \frac{1}{K_3}$.

- Assume property ii holds. As before, by homogeneity we may assume that $K_2 = 1$. Recalling the Taylor series expansion of the exponential function, we obtain

$$\begin{aligned}\mathbb{E}\left(e^{\lambda^2X^2}\right) &= \mathbb{E}\left(1 + \sum_{p=1}^{\infty} \frac{(\lambda^2X^2)^p}{p!}\right) \\ &= 1 + \sum_{p=1}^{\infty} \frac{\lambda^{2p}\mathbb{E}(X^{2p})}{p!}.\end{aligned}$$

Proof $\left(\mathbb{E}\left(|X|^p\right)\right)^{1/p} \le K_2\sqrt{p}$ **for all** $p \ge 1$ **implies** $\mathbb{E}\left(e^{\lambda^2 X^2}\right) \le \exp\left(K_3^2 \lambda^2\right), \quad \forall |\lambda| \le \frac{1}{K_3}$**.**

- Assume property ii holds. As before, by homogeneity we may assume that $K_2 = 1$. Recalling the Taylor series expansion of the exponential function, we obtain

$$\begin{aligned}\mathbb{E}\left(e^{\lambda^2 X^2}\right) &= \mathbb{E}\left(1 + \sum_{p=1}^{\infty} \frac{(\lambda^2 X^2)^p}{p!}\right) \\ &= 1 + \sum_{p=1}^{\infty} \frac{\lambda^{2p}\mathbb{E}(X^{2p})}{p!}.\end{aligned}$$

- Property 2 guarantees that $\mathbb{E}(X^{2p}) \le (2p)^p$,

Proof $\left(\mathbb{E}\left(|X|^p\right)\right)^{1/p} \le K_2\sqrt{p}$ **for all** $p \ge 1$ **implies** $\mathbb{E}\left(e^{\lambda^2 X^2}\right) \le \exp\left(K_3^2\lambda^2\right), \quad \forall|\lambda| \le \frac{1}{K_3}$**.**

- Assume property ii holds. As before, by homogeneity we may assume that $K_2 = 1$. Recalling the Taylor series expansion of the exponential function, we obtain

$$\begin{aligned}\mathbb{E}\left(e^{\lambda^2 X^2}\right) &= \mathbb{E}\left(1 + \sum_{p=1}^{\infty} \frac{(\lambda^2 X^2)^p}{p!}\right) \\ &= 1 + \sum_{p=1}^{\infty} \frac{\lambda^{2p}\mathbb{E}(X^{2p})}{p!}.\end{aligned}$$

- Property 2 guarantees that $\mathbb{E}(X^{2p}) \le (2p)^p$, while Stirling's approximation yields $p! \ge (p/e)^p$.

Proof $\left(\mathbb{E}\left(|X|^p\right)\right)^{1/p} \leq K_2\sqrt{p}$ **for all** $p \geq 1$ **implies** $\mathbb{E}\left(e^{\lambda^2 X^2}\right) \leq \exp\left(K_3^2\lambda^2\right), \quad \forall|\lambda| \leq \frac{1}{K_3}$.

- Assume property ii holds. As before, by homogeneity we may assume that $K_2 = 1$. Recalling the Taylor series expansion of the exponential function, we obtain

$$\begin{aligned}\mathbb{E}\left(e^{\lambda^2 X^2}\right) &= \mathbb{E}\left(1+\sum_{p=1}^{\infty}\frac{(\lambda^2X^2)^p}{p!}\right)\\ &= 1+\sum_{p=1}^{\infty}\frac{\lambda^{2p}\mathbb{E}(X^{2p})}{p!}.\end{aligned}$$

- Property 2 guarantees that $\mathbb{E}(X^{2p}) \leq (2p)^p$, while Stirling's approximation yields $p! \geq (p/e)^p$. Hence,

$$\mathbb{E}\left(e^{\lambda^2 X^2}\right) \leq 1+\sum_{p=1}^{\infty}\frac{(2\lambda^2 p)^p}{(p/e)^p}$$

Proof $(\mathbb{E}(|X|^p))^{1/p} \leq K_2\sqrt{p}$ **for all** $p \geq 1$ **implies** $\mathbb{E}\left(e^{\lambda^2 X^2}\right) \leq \exp\left(K_3^2\lambda^2\right), \quad \forall|\lambda| \leq \frac{1}{K_3}$**.**

- Assume property ii holds. As before, by homogeneity we may assume that $K_2 = 1$. Recalling the Taylor series expansion of the exponential function, we obtain

$$\begin{aligned}\mathbb{E}\left(e^{\lambda^2 X^2}\right) &= \mathbb{E}\left(1 + \sum_{p=1}^{\infty} \frac{(\lambda^2 X^2)^p}{p!}\right) \\ &= 1 + \sum_{p=1}^{\infty} \frac{\lambda^{2p}\mathbb{E}(X^{2p})}{p!}.\end{aligned}$$

- Property 2 guarantees that $\mathbb{E}(X^{2p}) \leq (2p)^p$, while Stirling's approximation yields $p! \geq (p/e)^p$. Hence,

$$\begin{aligned}\mathbb{E}\left(e^{\lambda^2 X^2}\right) &\leq 1 + \sum_{p=1}^{\infty} \frac{(2\lambda^2 p)^p}{(p/e)^p} \\ &= 1 + \sum_{p=1}^{\infty}(2e\lambda^2)^p\end{aligned}$$

Proof $\left(\mathbb{E}\left(|X|^p\right)\right)^{1/p} \le K_2\sqrt{p}$ **for all** $p \ge 1$ **implies** $\mathbb{E}\left(e^{\lambda^2 X^2}\right) \le \exp\left(K_3^2\lambda^2\right), \quad \forall|\lambda| \le \frac{1}{K_3}$**.**

- Assume property ii holds. As before, by homogeneity we may assume that $K_2 = 1$. Recalling the Taylor series expansion of the exponential function, we obtain

$$\begin{aligned}\mathbb{E}\left(e^{\lambda^2 X^2}\right) &= \mathbb{E}\left(1 + \sum_{p=1}^{\infty} \frac{(\lambda^2 X^2)^p}{p!}\right) \\ &= 1 + \sum_{p=1}^{\infty} \frac{\lambda^{2p}\mathbb{E}(X^{2p})}{p!}.\end{aligned}$$

- Property 2 guarantees that $\mathbb{E}(X^{2p}) \le (2p)^p$, while Stirling's approximation yields $p! \ge (p/e)^p$. Hence,

$$\begin{aligned}\mathbb{E}\left(e^{\lambda^2 X^2}\right) &\le 1 + \sum_{p=1}^{\infty} \frac{(2\lambda^2 p)^p}{(p/e)^p} \\ &= 1 + \sum_{p=1}^{\infty} (2e\lambda^2)^p \\ &= \frac{1}{1-2e\lambda^2}\end{aligned}$$

Proof $\left(\mathbb{E}\left(|X|^p\right)\right)^{1/p} \leq K_2\sqrt{p}$ **for all** $p \geq 1$ **implies** $\mathbb{E}\left(e^{\lambda^2 X^2}\right) \leq \exp\left(K_3^2\lambda^2\right), \quad \forall|\lambda| \leq \frac{1}{K_3}$**.**

- Assume property ii holds. As before, by homogeneity we may assume that $K_2 = 1$. Recalling the Taylor series expansion of the exponential function, we obtain

$$\begin{aligned}\mathbb{E}\left(e^{\lambda^2 X^2}\right) &= \mathbb{E}\left(1 + \sum_{p=1}^{\infty} \frac{(\lambda^2 X^2)^p}{p!}\right) \\ &= 1 + \sum_{p=1}^{\infty} \frac{\lambda^{2p}\mathbb{E}(X^{2p})}{p!}.\end{aligned}$$

- Property 2 guarantees that $\mathbb{E}(X^{2p}) \leq (2p)^p$, while Stirling's approximation yields $p! \geq (p/e)^p$. Hence,

$$\begin{aligned}\mathbb{E}\left(e^{\lambda^2 X^2}\right) &\leq 1 + \sum_{p=1}^{\infty} \frac{(2\lambda^2 p)^p}{(p/e)^p} \\ &= 1 + \sum_{p=1}^{\infty}(2e\lambda^2)^p \\ &= \frac{1}{1-2e\lambda^2}\end{aligned}$$

provided that $2e\lambda^2 < 1$.

Proof $\left(\mathbb{E}\left(|X|^p\right)\right)^{1/p} \leq K_2\sqrt{p}$ **for all** $p \geq 1$ **implies** $\mathbb{E}\left(e^{\lambda^2 X^2}\right) \leq \exp\left(K_3^2\lambda^2\right), \quad \forall|\lambda| \leq \frac{1}{K_3}$**.**

- We have shown that

$$\mathbb{E}\left(e^{\lambda^2 X^2}\right) \leq \frac{1}{1 - 2e\lambda^2}$$

provided that $2e\lambda^2 < 1$.

Proof $(\mathbb{E}(|X|^p))^{1/p} \le K_2\sqrt{p}$ **for all** $p \ge 1$ **implies** $\mathbb{E}\left(e^{\lambda^2 X^2}\right) \le \exp\left(K_3^2\lambda^2\right), \quad \forall|\lambda| \le \frac{1}{K_3}$.

- We have shown that

$$\mathbb{E}\left(e^{\lambda^2 X^2}\right) \le \frac{1}{1 - 2e\lambda^2}$$

  provided that $2e\lambda^2 < 1$.
- Using the inequality $1/(1-x) \le e^{2x}$ for all $x \in [0, 1/2]$, we obtain

$$\mathbb{E}\left(e^{\lambda^2 X^2}\right) \le \exp(4e\lambda^2)$$

  for all λ such that $2e\lambda^2 < 1/2$ or $|\lambda| < 1/(2\sqrt{e})$.

Sub-Gaussian norm

- The sub-Gaussian norm is defined as

$$\|X\|_{\psi_2} := \inf \left\{ t > 0 : \mathbb{E}\left(\exp\left(X^2/t^2\right)\right) \leq 2 \right\}$$

Sub-Gaussian norm

- The sub-Gaussian norm is defined as

$$\|X\|_{\psi_2} := \inf\left\{t > 0 : \mathbb{E}\left(\exp\left(X^2/t^2\right)\right) \leq 2\right\}$$

- X is sub-Gaussian if and only if $\|X\|_{\psi_2} < \infty$.

Sub-Gaussian norm

- The sub-Gaussian norm is defined as

$$\|X\|_{\psi_2} := \inf\left\{t > 0 : \mathbb{E}\left(\exp\left(X^2/t^2\right)\right) \leq 2\right\}$$

- X is sub-Gaussian if and only if $\|X\|_{\psi_2} < \infty$. This follows from Property 4.

Sub-Gaussian norm

- The sub-Gaussian norm is defined as

$$\|X\|_{\psi_2} := \inf\left\{t > 0 : \mathbb{E}\left(\exp\left(X^2/t^2\right)\right) \leq 2\right\}$$

- X is sub-Gaussian if and only if $\|X\|_{\psi_2} < \infty$. This follows from Property 4.
- $\|\cdot\|_{\psi_2}$ is a proper norm on the space of sub-Gaussian random variables.

- Every sub-Gaussian variable satisfies the following bounds:

$$
\begin{aligned}
\mathbb{P}\left(|X| \geq t\right) &\leq 2\exp\left(-ct^2/\|X\|_{\psi_2}^2\right), \quad \forall t \geq 0. \\
\|X\|_p &\leq C\|X\|_{\psi_2}\sqrt{p}, \ \ \forall p \geq 1. \\
\mathbb{E}\left(\exp\left(X^2/\|X\|_{\psi_2}^2\right)\right) &\leq 2. \qquad (X \text{ nonzero }).
\end{aligned}
$$

- Every sub-Gaussian variable satisfies the following bounds:

$$\begin{aligned}
\mathbb{P}\left(|X| \geq t\right) &\leq 2\exp\left(-ct^2/\|X\|_{\psi_2}^2\right), \quad \forall t \geq 0. \\
\|X\|_p &\leq C\|X\|_{\psi_2}\sqrt{p}, \quad \forall p \geq 1. \\
\mathbb{E}\left(\exp\left(X^2/\|X\|_{\psi_2}^2\right)\right) &\leq 2. \qquad (X \text{ nonzero }).
\end{aligned}$$

- When $\mathbb{E}(X) = 0$, $\mathbb{E}\left(\exp\left(\lambda X\right)\right) \leq \exp\left(C_1\lambda^2\|X\|_{\psi_2}^2\right)$, for all $\lambda \geq 0$, and some universal constants $C, C_1, c > 0$.

Some consequences

- Recall what a universal/numerical/absolute constant means.

Some consequences

- Recall what a universal/numerical/absolute constant means.
- Sub-Gaussian norm is within a constant factor of the sub-Gaussian parameter σ : for numerical constant $c_1, c_2 > 0$:

$$c_1\|X\|_{\psi_2} \leq \sigma(X) \leq c_2\|X\|_{\psi_2}.$$

Some consequences

- Recall what a universal/numerical/absolute constant means.
- Sub-Gaussian norm is within a constant factor of the sub-Gaussian parameter σ : for numerical constant $c_1, c_2 > 0$:

$$c_1 \|X\|_{\psi_2} \le \sigma(X) \le c_2 \|X\|_{\psi_2}.$$

- Easy to see that $\|X\|_{\psi_2} \lesssim \|X\|_\infty$ (Bounded variables are sub-Gaussian). Here, $a \lesssim b$ means $a \le Cb$ for some constant $C > 0$.

Some consequences

- Recall what a universal/numerical/absolute constant means.
- Sub-Gaussian norm is within a constant factor of the sub-Gaussian parameter σ : for numerical constant $c_1, c_2 > 0$:

$$c_1 \|X\|_{\psi_2} \le \sigma(X) \le c_2 \|X\|_{\psi_2}.$$

- Easy to see that $\|X\|_{\psi_2} \lesssim \|X\|_\infty$ (Bounded variables are sub-Gaussian). Here, $a \lesssim b$ means $a \le Cb$ for some constant $C > 0$.
- **Lemma.** If X is sub-Gaussian then $X - \mathbb{E}(X)$ is sub-Gaussian too and

$$\|X - \mathbb{E}(X)\|_{\psi_2} \le C \|X\|_{\psi_2}$$

where $C > 0$ is a universal constant.

- **Proposition.** Assume that $\{X_i\}$ are independent, zero-mean sub-Gaussian random variables. Then $\sum_{i=1}^n X_i$ is also sub-Gaussian and

$$\|\sum_{i=1}^n X_i\|_{\psi_2}^2 \le C\sum_{i=1}^n \|X_i\|_{\psi_2}^2,$$

for some positive constant $C > 0$.

- **Proposition.** Assume that $\{X_i\}$ are independent, zero-mean sub-Gaussian random variables. Then $\sum_{i=1}^n X_i$ is also sub-Gaussian and

$$\|\sum_{i=1}^n X_i\|_{\psi_2}^2 \le C \sum_{i=1}^n \|X_i\|_{\psi_2}^2,$$

  for some positive constant $C > 0$.

Proof.

- Let X_i be sub-Gaussian with parameter σ_i. Then, we know that $\sum_{i=1}^n X_i$ is sub-Gaussian with parameter $\sqrt{\sum_{i=1}^n \sigma_i^2}$.

- **Proposition.** Assume that $\{X_i\}$ are independent, zero-mean sub-Gaussian random variables. Then $\sum_{i=1}^n X_i$ is also sub-Gaussian and

$$\|\sum_{i=1}^n X_i\|_{\psi_2}^2 \le C\sum_{i=1}^n \|X_i\|_{\psi_2}^2,$$

  for some positive constant $C > 0$.

Proof.

- Let X_i be sub-Gaussian with parameter σ_i. Then, we know that $\sum_{i=1}^n X_i$ is sub-Gaussian with parameter $\sqrt{\sum_{i=1}^n \sigma_i^2}$. Hence,

$$\|\sum_{i=1}^n X_i\|_{\psi_2}^2 \lesssim \sum_{i=1}^n \sigma_i^2 \lesssim \sum_{i=1}^n \|X_i\|_{\psi_2}^2.$$

$\square$

- **Theorem.** Assume that $\{X_i\}$ are independent, zero-mean sub-Gaussian random variables. Then

$$\mathbb{P}\left(|\sum_{i=1}^{n} X_i| \geq t \right) \leq 2 \exp\left(-\frac{ct^2}{\sum_{i=1}^{n} \|X_i\|_{\psi_2}^2} \right), \quad t \geq 0,$$

where c is auniversal constant.

.

Sub-exponential concentration

- **Definition 2.** A zero-mean random variable X is sub-exponential if for some $\nu, \alpha > 0$,

$$\mathbb{E}\left(\exp(\lambda X)\right) \leq e^{\nu^2 \lambda^2 / 2}, \quad \forall |\lambda| < \frac{1}{\alpha}. \tag{2}$$

A general random variable is sub-exponential if $X - \mathbb{E}(X)$ is sub-Exponential.

Sub-exponential concentration

- **Definition 2.** A zero-mean random variable X is sub-exponential if for some $\nu, \alpha > 0$,

$$\mathbb{E}\left(\exp(\lambda X)\right) \leq e^{\nu^2\lambda^2/2}, \quad \forall|\lambda| < \frac{1}{\alpha}. \tag{2}$$

  A general random variable is sub-exponential if $X - \mathbb{E}(X)$ is sub-Exponential.
- If $Z \sim N(0,1)$ then $Z^2 - 1$ is sub-Exponential:

$$\mathbb{E}\left(e^{\lambda(Z^2-1)}\right) = \begin{cases} \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} & \lambda < 1/2 \\ \infty & \lambda > 1/2. \end{cases}$$

Sub-exponential concentration

- **Definition 2.** A zero-mean random variable X is sub-exponential if for some $\nu, \alpha > 0$,

$$\mathbb{E}\left(\exp(\lambda X)\right) \leq e^{\nu^2\lambda^2/2}, \quad \forall |\lambda| < \frac{1}{\alpha}. \tag{2}$$

  A general random variable is sub-exponential if $X - \mathbb{E}(X)$ is sub-Exponential.
- If $Z \sim N(0,1)$ then $Z^2 - 1$ is sub-Exponential:

$$\mathbb{E}\left(e^{\lambda(Z^2-1)}\right) = \begin{cases} \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} & \lambda < 1/2 \\ \infty & \lambda > 1/2. \end{cases}$$

- We have $\mathbb{E}(e^{\lambda(Z^2-1)}) \leq e^{4\lambda^2/2}$, $\lambda < 1/4$.

Sub-exponential concentration

- **Definition 2.** A zero-mean random variable X is sub-exponential if for some $\nu, \alpha > 0$,

$$\mathbb{E}\left(\exp(\lambda X)\right) \leq e^{\nu^2\lambda^2/2}, \quad \forall|\lambda| < \frac{1}{\alpha}. \tag{2}$$

  A general random variable is sub-exponential if $X - \mathbb{E}(X)$ is sub-Exponential.
- If $Z \sim N(0,1)$ then $Z^2 - 1$ is sub-Exponential:

$$\mathbb{E}\left(e^{\lambda(Z^2-1)}\right) = \begin{cases} \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} & \lambda < 1/2 \\ \infty & \lambda > 1/2. \end{cases}$$

- We have $\mathbb{E}(e^{\lambda(Z^2-1)}) \leq e^{4\lambda^2/2}$, $\lambda < 1/4$.
- Hence $Z^2 - 1$ is sub-exponential with parameters $(2, 4)$.

Sub-exponential concentration

- **Definition 2.** A zero-mean random variable X is sub-exponential if for some $\nu, \alpha > 0$,

$$\mathbb{E}\left(\exp(\lambda X)\right) \leq e^{\nu^2\lambda^2/2}, \quad \forall |\lambda| < \frac{1}{\alpha}. \tag{2}$$

  A general random variable is sub-exponential if $X - \mathbb{E}(X)$ is sub-Exponential.
- If $Z \sim N(0, 1)$ then $Z^2 - 1$ is sub-Exponential:

$$\mathbb{E}\left(e^{\lambda(Z^2-1)}\right) = \begin{cases} \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} & \lambda < 1/2 \\ \infty & \lambda > 1/2. \end{cases}$$

- We have $\mathbb{E}(e^{\lambda(Z^2-1)}) \leq e^{4\lambda^2/2}$, $\lambda < 1/4$.
- Hence $Z^2 - 1$ is sub-exponential with parameters $(2, 4)$.
- Tails of $Z^2 - 1$ are heavier than a Gaussian.

- **Proposition 4.** Assume that X is zero-mean sub-exponential satisfying (2). Then

$$\mathbb{P}(X \geq t) \leq \exp\left(-\frac{1}{2}\min\left\{\frac{t^2}{\nu}, \frac{t}{\alpha}\right\}\right), \quad \forall t \geq 0,$$

  Same bound holds with X replaced with $-X$.

- **Proposition 4.** Assume that X is zero-mean sub-exponential satisfying (2). Then

$$\mathbb{P}(X \geq t) \leq \exp\left(-\frac{1}{2}\min\left\{\frac{t^2}{\nu}, \frac{t}{\alpha}\right\}\right), \quad \forall t \geq 0,$$

Same bound holds with X replaced with $-X$.
Proof. Chernoff bound

$$\begin{aligned}\mathbb{P}(X \geq t) &\leq \inf_{0<\lambda<1/\alpha} \left[e^{-\lambda t}\mathbb{E}\left(e^{\lambda X}\right)\right] \\ &\leq \inf_{0<\lambda<\frac{1}{\alpha}} \exp\left(-\lambda t + \frac{\lambda^2\nu^2}{2}\right).\end{aligned}$$

- **Proposition 4.** Assume that X is zero-mean sub-exponential satisfying (2). Then

$$\mathbb{P}(X \geq t) \leq \exp\left(-\frac{1}{2}\min\left\{\frac{t^2}{\nu}, \frac{t}{\alpha}\right\}\right), \quad \forall t \geq 0,$$

  Same bound holds with X replaced with $-X$.
  Proof. Chernoff bound

$$\begin{aligned}\mathbb{P}(X \geq t) &\leq \inf_{0<\lambda<1/\alpha} \left[e^{-\lambda t}\mathbb{E}\left(e^{\lambda X}\right)\right] \\ &\leq \inf_{0<\lambda<\frac{1}{\alpha}} \exp\left(-\lambda t + \tfrac{\lambda^2\nu^2}{2}\right).\end{aligned}$$

- Let $f(\lambda) = -\lambda t + \frac{\lambda^2\nu^2}{2}$

- **Proposition 4.** Assume that X is zero-mean sub-exponential satisfying (2). Then

$$\mathbb{P}(X \geq t) \leq \exp\left(-\frac{1}{2}\min\left\{\frac{t^2}{\nu}, \frac{t}{\alpha}\right\}\right), \quad \forall t \geq 0,$$

  Same bound holds with X replaced with $-X$.
  Proof. Chernoff bound

$$\begin{aligned} \mathbb{P}(X \geq t) &\leq \inf_{0<\lambda<1/\alpha} \left[e^{-\lambda t}\mathbb{E}\left(e^{\lambda X}\right)\right] \\ &\leq \inf_{0<\lambda<\frac{1}{\alpha}} \exp\left(-\lambda t + \frac{\lambda^2\nu^2}{2}\right). \end{aligned}$$

- Let $f(\lambda) = -\lambda t + \frac{\lambda^2\nu^2}{2}$
- Minimizer of f over λ is $\lambda = t/\nu^2$.

- Hence minimizer of f over $[0, 1/\alpha]$ is

$$\lambda^* = \begin{cases} t/\nu^2 & t/\nu^2 < 1/\alpha \\ 1/\alpha & t/\nu^2 \geq 1/\alpha. \end{cases}$$

- Hence minimizer of f over $[0, 1/\alpha]$ is

$$\lambda^* = \begin{cases} t/\nu^2 & t/\nu^2 < 1/\alpha \\ 1/\alpha & t/\nu^2 \geq 1/\alpha. \end{cases}$$

- And the minimum is

$$f(\lambda^*) = \begin{cases} -\frac{t^2}{2\nu^2} & t < \frac{\nu^2}{\alpha} \\ -\frac{t}{\alpha} + \frac{\nu^2}{2\alpha^2} \leq -\frac{t}{2\alpha} & t \geq \frac{\nu^2}{\alpha}. \end{cases}$$

- Hence minimizer of f over $[0, 1/\alpha]$ is

$$\lambda^* = \begin{cases} t/\nu^2 & t/\nu^2 < 1/\alpha \\ 1/\alpha & t/\nu^2 \geq 1/\alpha. \end{cases}$$

- And the minimum is

$$f(\lambda^*) = \begin{cases} -\frac{t^2}{2\nu^2} & t < \frac{\nu^2}{\alpha} \\ -\frac{t}{\alpha} + \frac{\nu^2}{2\alpha^2} \leq -\frac{t}{2\alpha} & t \geq \frac{\nu^2}{\alpha}. \end{cases}$$

- Thus,

$$f(\lambda^*) \leq \max\{-\frac{t^2}{2\nu^2}, -\frac{t}{2\alpha}\} = -\frac{1}{2}\min\left\{\frac{t^2}{\nu^2}, \frac{t}{\alpha}\right\}$$

Bernstein inequality for sub-exponential random variables

- **Theorem 4.(Bernstein)** Assume that $\{X_i\}$ are independent, zero-mean sub-exponential RVs with parameters (α_i, ν_i). Let $\nu = \sqrt{\sum_{i=1}^n \nu_i^2}$ and $\alpha = \max_i \alpha_i$. Then $X = \sum_{i=1}^n X_i$ is sub-Exponential with parameters (ν, α) and

$$\mathbb{P}(X \geq t) \leq \exp\left(-\frac{1}{2}\min\left\{\frac{t^2}{\nu^2}, \frac{t}{\alpha}\right\}\right).$$

- **Proof.** We have

$$\mathbb{E}\left(e^{\lambda X_i}\right) \le e^{\lambda^2 \nu_i^2/2}, \qquad \forall |\lambda| < \frac{1}{\alpha} \le \frac{1}{\alpha_i}.$$

- **Proof.** We have

$$\mathbb{E}\left(e^{\lambda X_i}\right) \le e^{\lambda^2 \nu_i^2/2}, \qquad \forall|\lambda| < \frac{1}{\alpha} \le \frac{1}{\alpha_i}.$$

- Let $S_n = \sum_{i=1}^n X_i$. By independence

$$\mathbb{E}\left(e^{\lambda S_n}\right) = \prod_{i=1}^n \mathbb{E}\left(e^{\lambda X_i}\right) \le e^{\lambda^2 \sum_{i=1}^n \nu_i^2/2}, \quad \forall|\lambda| < \frac{1}{\alpha}$$

- **Proof.** We have

$$\mathbb{E}\left(e^{\lambda X_i}\right) \leq e^{\lambda^2 \nu_i^2/2}, \qquad \forall |\lambda| < \frac{1}{\alpha} \leq \frac{1}{\alpha_i}.$$

- Let $S_n = \sum_{i=1}^n X_i$. By independence

$$\mathbb{E}\left(e^{\lambda S_n}\right) = \prod_{i=1}^n \mathbb{E}\left(e^{\lambda X_i}\right) \leq e^{\lambda^2 \sum_{i=1}^n \nu_i^2/2}, \quad \forall |\lambda| < \frac{1}{\alpha}$$

- The tail bound follows from Proposition 4.

Equivalent characterizations of sub-exponential random variables

For a RV X, the following are equivalent: (HDP, Prop. 2.7.1)

- The tails of X satisfy

$$\mathbb{P}\left(|X| \geq t\right) \leq 2\exp\left(-\frac{t}{K_1}\right), \quad \forall t \geq 0.$$

Equivalent characterizations of sub-exponential random variables

For a RV X, the following are equivalent: (HDP, Prop. 2.7.1)

- The tails of X satisfy

$$\mathbb{P}\left(|X| \geq t\right) \leq 2\exp\left(-\frac{t}{K_1}\right), \quad \forall t \geq 0.$$

- The moments of X satisfy

$$\|X\|_p = \left(\mathbb{E}\left(|X|^p\right)\right)^{1/p} \leq K_2 p, \quad \forall \geq 1.$$

Equivalent characterizations of sub-exponential random variables

For a RV X, the following are equivalent: (HDP, Prop. 2.7.1)

- The tails of X satisfy
$$\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t}{K_1}\right), \quad \forall t \geq 0.$$
- The moments of X satisfy
$$\|X\|_p = \left(\mathbb{E}(|X|^p)\right)^{1/p} \leq K_2 p, \quad \forall \geq 1.$$
- The MGF of $|X|$ is bounded at some point,
$$\mathbb{E}(\exp(|X|/K_4)) \leq 2.$$

Equivalent characterizations of sub-exponential random variables

For a RV X, the following are equivalent: (HDP, Prop. 2.7.1)

- The tails of X satisfy

$$\mathbb{P}\left(|X| \geq t\right) \leq 2\exp\left(-\frac{t}{K_1}\right), \quad \forall t \geq 0.$$

- The moments of X satisfy

$$\|X\|_p = \left(\mathbb{E}\left(|X|^p\right)\right)^{1/p} \leq K_2 p, \quad \forall \geq 1.$$

- The MGF of $|X|$ is bounded at some point,

$$\mathbb{E}\left(\exp\left(|X|/K_4\right)\right) \leq 2.$$

- Assuming $\mathbb{E}(X) = 0$, the above are equivalent to:

$$\mathbb{E}\left(\exp\left(\lambda X\right)\right) \leq \exp\left(K_5^2\lambda^2\right), \quad \forall|\lambda| \leq \frac{1}{K_5}.$$

Sub-exponential norm

- The sub-exponential norm is defined as

$$\|X\|_{\psi_1} = \inf\left\{t > 0 : \mathbb{E}\left(\exp\left(|X|/t\right)\right) \le 2\right\}$$

Sub-exponential norm

- The sub-exponential norm is defined as

$$\|X\|_{\psi_1} = \inf\{t > 0 : \mathbb{E}(\exp(|X|/t)) \leq 2\}$$

- X is sub-exponential if and only if $\|X\|_{\psi_1} < \infty$.

Sub-exponential norm

- The sub-exponential norm is defined as

$$\|X\|_{\psi_1} = \inf\left\{t > 0 : \mathbb{E}\left(\exp\left(|X|/t\right)\right) \leq 2\right\}$$

- X is sub-exponential if and only if $\|X\|_{\psi_1} < \infty$.
- $\|X\|_{\psi_1}$ is a proper norm on the space of sub-exponential random variables.

Sub-exponential norm

- The sub-exponential norm is defined as

$$\|X\|_{\psi_1} = \inf\left\{t > 0 \,:\, \mathbb{E}\left(\exp\left(|X|/t\right)\right) \leq 2\right\}$$

- X is sub-exponential if and only if $\|X\|_{\psi_1} < \infty$.
- $\|X\|_{\psi_1}$ is a proper norm on the space of sub-exponential random variables.
- Every sub-exponential variable satisfies the following bounds:

$$\begin{aligned}
\mathbb{P}\left(|X| \geq t\right) &\leq 2\exp\left(-ct/\|X\|_{\psi_1}\right), \quad \forall t \geq 0. \\
\|X\|_p &\leq C\|X\|_{\psi_1} p, \quad \forall p \geq 1. \\
\mathbb{E}\left(\exp\left(|X|/\|X\|_{\psi_1}\right)\right) &\leq 2.
\end{aligned}$$

Sub-exponential norm

- The sub-exponential norm is defined as

$$\|X\|_{\psi_1} = \inf\{t > 0 : \mathbb{E}(\exp(|X|/t)) \leq 2\}$$

- X is sub-exponential if and only if $\|X\|_{\psi_1} < \infty$.
- $\|X\|_{\psi_1}$ is a proper norm on the space of sub-exponential random variables.
- Every sub-exponential variable satisfies the following bounds:

$$\begin{aligned}
\mathbb{P}(|X| \geq t) &\leq 2\exp\left(-ct/\|X\|_{\psi_1}\right), \quad \forall t \geq 0. \\
\|X\|_p &\leq C\|X\|_{\psi_1} p, \quad \forall p \geq 1. \\
\mathbb{E}\left(\exp\left(|X|/\|X\|_{\psi_1}\right)\right) &\leq 2.
\end{aligned}$$

- When $\mathbb{E}(X) = 0$, $\mathbb{E}(\exp(\lambda X)) \leq \exp\left(C\lambda^2\|X\|_{\psi_1}\right)$, for all $|\lambda| \leq 1/\|X\|_{\psi_1}$.
  for some universal constants $C, c > 0$.

- **Lemma 2** A random variable X is sub-Gaussian if and only if X^2 is sub-Exponential, in fact

$$\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2.$$

- **Lemma 2** A random variable X is sub-Gaussian if and only if X^2 is sub-Exponential, in fact

$$\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2.$$

- Proof: Immediate from definition.

- **Lemma 2** A random variable X is sub-Gaussian if and only if X^2 is sub-Exponential, in fact

$$\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2.$$

- Proof: Immediate from definition.
- **Lemma 3.** If X and Y are sub-Gaussian, then XY is sub-exponential, and

$$\|XY\|_{\psi_1} \leq \|X\|_{\psi_2}\|Y\|_{\psi_2}.$$

ℓ_∞ norm of sub-Gaussian vectors

- For any vector $X \in \mathbb{R}^n$, the ℓ_∞ norm is

$$\|X\|_\infty = \max_{i=1,\ldots,n} |X_i|.$$

ℓ_∞ norm of sub-Gaussian vectors

- For any vector $X \in \mathbb{R}^n$, the ℓ_∞ norm is

$$\|X\|_\infty = \max_{i=1,\ldots,n} |X_i|.$$

- Notice that

$$\frac{1}{\sqrt{n}}\|X\|_2 = \frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n} X_i^2} \leq \|X\|_\infty.$$

ℓ_∞ norm of sub-Gaussian vectors

- For any vector $X \in \mathbb{R}^n$, the ℓ_∞ norm is

$$\|X\|_\infty = \max_{i=1,\ldots,n} |X_i|.$$

- Notice that

$$\frac{1}{\sqrt{n}}\|X\|_2 = \frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^n X_i^2} \leq \|X\|_\infty.$$

- **Lemma 5.** Let $X = (X_1, \ldots, X_n)^\top \in \mathbb{R}^n$ be a random vector with zero-mean, sub-Gaussian coordinates X_i with parameter σ_i. Then, for any $\gamma > 0$,

$$\mathbb{P}\left(\|X\|_\infty \geq \sigma\sqrt{2(1+\gamma)\log n}\right) \leq 2n^{-\gamma}$$

where $\sigma = \max_i \sigma_i$.

Proof

- We have by Sub-Gaussianity that

$$\mathbb{P}\left(|X_i| \geq t\right) \leq \exp\left(-\frac{t^2}{2\sigma_i^2}\right) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

Proof

- We have by Sub-Gaussianity that

$$\mathbb{P}\left(|X_i| \geq t\right) \leq \exp\left(-\frac{t^2}{2\sigma_i^2}\right) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

- Hence,

$$\mathbb{P}\left(\max_i |X_i| \geq t\right) \quad \leq \quad \mathbb{P}\left(\text{at least one } i \text{ s.t } |X_i| \geq t\right)$$

Proof

- We have by Sub-Gaussianity that

$$\mathbb{P}\left(|X_i| \geq t\right) \leq \exp\left(-\frac{t^2}{2\sigma_i^2}\right) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

- Hence,

$$\begin{aligned} \mathbb{P}\left(\max_i |X_i| \geq t\right) &\leq \mathbb{P}\left(\text{at least one } i \text{ s.t } |X_i| \geq t\right) \\ &= \mathbb{P}\left(\bigcup_{i=1}^n \{|X_i| \geq t\}\right) \end{aligned}$$

Proof

- We have by Sub-Gaussianity that

$$\mathbb{P}\left(|X_i| \geq t\right) \leq \exp\left(-\frac{t^2}{2\sigma_i^2}\right) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

- Hence,

$$\begin{array}{rcl} \mathbb{P}\left(\max_i |X_i| \geq t\right) & \leq & \mathbb{P}\left(\text{at least one } i \text{ s.t } |X_i| \geq t\right) \\ & = & \mathbb{P}\left(\bigcup_{i=1}^n \{|X_i| \geq t\}\right) \\ & \leq & \sum_{i=1}^n \mathbb{P}\left(|X_i| \geq t\right) \end{array}$$

Proof

- We have by Sub-Gaussianity that

$$\mathbb{P}\left(|X_i| \geq t\right) \leq \exp\left(-\frac{t^2}{2\sigma_i^2}\right) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

- Hence,

$$\begin{array}{rcl} \mathbb{P}\left(\max_i |X_i| \geq t\right) & \leq & \mathbb{P}\left(\text{at least one } i \text{ s.t } |X_i| \geq t\right) \\ & = & \mathbb{P}\left(\bigcup_{i=1}^n \{|X_i| \geq t\}\right) \\ & \leq & \sum_{i=1}^n \mathbb{P}\left(|X_i| \geq t\right) \\ & \leq & \sum_{i=1}^n \exp\left(-\frac{t^2}{2\sigma^2}\right) \end{array}$$

Proof

- We have by Sub-Gaussianity that

$$\mathbb{P}\left(|X_i| \geq t\right) \leq \exp\left(-\frac{t^2}{2\sigma_i^2}\right) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

- Hence,

$$\begin{aligned}
\mathbb{P}\left(\max_i |X_i| \geq t\right) &\leq \mathbb{P}\left(\text{at least one } i \text{ s.t } |X_i| \geq t\right) \\
&= \mathbb{P}\left(\bigcup_{i=1}^n \{|X_i| \geq t\}\right) \\
&\leq \sum_{i=1}^n \mathbb{P}\left(|X_i| \geq t\right) \\
&\leq \sum_{i=1}^n \exp\left(-\frac{t^2}{2\sigma^2}\right) \\
&= n \exp\left(-\frac{t^2}{2\sigma^2}\right)
\end{aligned}$$

Proof

- We have by Sub-Gaussianity that

$$\mathbb{P}\left(|X_i| \geq t\right) \;\leq\; \exp\left(-\frac{t^2}{2\sigma_i^2}\right) \;\leq\; \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

- Hence,

$$\begin{array}{rcl} \mathbb{P}\left(\max_i |X_i| \geq t\right) & \leq & \mathbb{P}\left(\text{at least one } i \text{ s.t } |X_i| \geq t\right) \\ & = & \mathbb{P}\left(\bigcup_{i=1}^n \{|X_i| \geq t\}\right) \\ & \leq & \sum_{i=1}^n \mathbb{P}\left(|X_i| \geq t\right) \\ & \leq & \sum_{i=1}^n \exp\left(-\frac{t^2}{2\sigma^2}\right) \\ & = & n \exp\left(-\frac{t^2}{2\sigma^2}\right) \\ & = & 2n^{-\gamma} \end{array}$$

taking $t = \sqrt{2\sigma^2(1+\gamma)\log n}$.

- **Theorem 6.** Assume that $\{X_i\}$ are zero-mean random variables, sub-Gaussian with parameter σ. Then

$$\mathbb{E}\left(\max_{i=1,\ldots,n} X_i\right) \leq \sqrt{2\sigma^2 \log n}, \quad n \geq 1.$$

- **Theorem 6.** Assume that $\{X_i\}$ are zero-mean random variables, sub-Gaussian with parameter σ. Then

$$\mathbb{E}\left(\max_{i=1,\ldots,n} X_i\right) \leq \sqrt{2\sigma^2 \log n}, \quad n \geq 1.$$

- **Proof.** For any $s > 0$,

$$\mathbb{E}\left(\max_{i=1,\ldots,n} X_i\right) = \frac{1}{s}\mathbb{E}\left(\log e^{s \max_{1\leq i\leq n} X_i}\right)$$

- **Theorem 6.** Assume that $\{X_i\}$ are zero-mean random variables, sub-Gaussian with parameter σ. Then

$$\mathbb{E}\left(\max_{i=1,\ldots,n} X_i\right) \le \sqrt{2\sigma^2 \log n}, \quad n \ge 1.$$

- **Proof.** For any $s > 0$,

$$\begin{aligned}
\mathbb{E}\left(\max_{i=1,\ldots,n} X_i\right) &= \frac{1}{s}\mathbb{E}\left(\log e^{s \max_{1\le i\le n} X_i}\right) \\
&\le \frac{1}{s}\log\left(\mathbb{E}\left(e^{s \max_{1\le i\le n} X_i}\right)\right)
\end{aligned}$$

- **Theorem 6.** Assume that $\{X_i\}$ are zero-mean random variables, sub-Gaussian with parameter σ. Then

$$\mathbb{E}\left(\max_{i=1,\dots,n} X_i\right) \leq \sqrt{2\sigma^2 \log n}, \quad n \geq 1.$$

- **Proof.** For any $s > 0$,

$$\begin{aligned}
\mathbb{E}\left(\max_{i=1,\dots,n} X_i\right) &= \tfrac{1}{s}\mathbb{E}\left(\log e^{s \max_{1\leq i\leq n} X_i}\right) \\
&\leq \tfrac{1}{s}\log\left(\mathbb{E}\left(e^{s \max_{1\leq i\leq n} X_i}\right)\right) \\
&= \tfrac{1}{s}\log\left(\mathbb{E}\left(\max_{1\leq i\leq n} e^{sX_i}\right)\right)
\end{aligned}$$

- **Theorem 6.** Assume that $\{X_i\}$ are zero-mean random variables, sub-Gaussian with parameter σ. Then

$$\mathbb{E}\left(\max_{i=1,\ldots,n} X_i\right) \le \sqrt{2\sigma^2 \log n}, \quad n \ge 1.$$

- **Proof.** For any $s > 0$,

$$\begin{aligned}
\mathbb{E}\left(\max_{i=1,\ldots,n} X_i\right) &= \tfrac{1}{s}\mathbb{E}\left(\log e^{s \max_{1\le i\le n} X_i}\right) \\
&\le \tfrac{1}{s}\log\left(\mathbb{E}\left(e^{s \max_{1\le i\le n} X_i}\right)\right) \\
&= \tfrac{1}{s}\log\left(\mathbb{E}\left(\max_{1\le i\le n} e^{sX_i}\right)\right) \\
&\le \tfrac{1}{s}\log\left(\mathbb{E}\left(\textstyle\sum_{i=1}^{n} e^{sX_i}\right)\right)
\end{aligned}$$

- **Theorem 6.** Assume that $\{X_i\}$ are zero-mean random variables, sub-Gaussian with parameter σ. Then

$$\mathbb{E}\left(\max_{i=1,\ldots,n} X_i\right) \leq \sqrt{2\sigma^2 \log n}, \quad n \geq 1.$$

- **Proof.** For any $s > 0$,

$$\begin{aligned}
\mathbb{E}\left(\max_{i=1,\ldots,n} X_i\right) &= \tfrac{1}{s}\mathbb{E}\left(\log e^{s \max_{1\leq i\leq n} X_i}\right) \\
&\leq \tfrac{1}{s}\log\left(\mathbb{E}\left(e^{s \max_{1\leq i\leq n} X_i}\right)\right) \\
&= \tfrac{1}{s}\log\left(\mathbb{E}\left(\max_{1\leq i\leq n} e^{sX_i}\right)\right) \\
&\leq \tfrac{1}{s}\log\left(\mathbb{E}\left(\textstyle\sum_{i=1}^{n} e^{sX_i}\right)\right) \\
&= \tfrac{1}{s}\log\left(\textstyle\sum_{i=1}^{n} \mathbb{E}\left(e^{sX_i}\right)\right)
\end{aligned}$$

- **Theorem 6.** Assume that $\{X_i\}$ are zero-mean random variables, sub-Gaussian with parameter σ. Then

$$\mathbb{E}\left(\max_{i=1,\ldots,n} X_i\right) \leq \sqrt{2\sigma^2 \log n}, \quad n \geq 1.$$

- **Proof.** For any $s > 0$,

$$\begin{aligned}
\mathbb{E}\left(\max_{i=1,\ldots,n} X_i\right) &= \tfrac{1}{s}\mathbb{E}\left(\log e^{s \max_{1\leq i\leq n} X_i}\right) \\
&\leq \tfrac{1}{s}\log\left(\mathbb{E}\left(e^{s \max_{1\leq i\leq n} X_i}\right)\right) \\
&= \tfrac{1}{s}\log\left(\mathbb{E}\left(\max_{1\leq i\leq n} e^{sX_i}\right)\right) \\
&\leq \tfrac{1}{s}\log\left(\mathbb{E}\left(\textstyle\sum_{i=1}^n e^{sX_i}\right)\right) \\
&= \tfrac{1}{s}\log\left(\textstyle\sum_{i=1}^n \mathbb{E}\left(e^{sX_i}\right)\right) \\
&\leq \tfrac{1}{s}\log\left(\textstyle\sum_{i=1}^n e^{\frac{\sigma^2 s^2}{2}}\right)
\end{aligned}$$

- **Theorem 6.** Assume that $\{X_i\}$ are zero-mean random variables, sub-Gaussian with parameter σ. Then

$$\mathbb{E}\left(\max_{i=1,\dots,n} X_i\right) \le \sqrt{2\sigma^2 \log n}, \quad n \ge 1.$$

- **Proof.** For any $s > 0$,

$$\begin{aligned}
\mathbb{E}\left(\max_{i=1,\dots,n} X_i\right) &= \tfrac{1}{s}\mathbb{E}\left(\log e^{s \max_{1\le i\le n} X_i}\right) \\
&\le \tfrac{1}{s}\log\left(\mathbb{E}\left(e^{s \max_{1\le i\le n} X_i}\right)\right) \\
&= \tfrac{1}{s}\log\left(\mathbb{E}\left(\max_{1\le i\le n} e^{sX_i}\right)\right) \\
&\le \tfrac{1}{s}\log\left(\mathbb{E}\left(\textstyle\sum_{i=1}^n e^{sX_i}\right)\right) \\
&= \tfrac{1}{s}\log\left(\textstyle\sum_{i=1}^n \mathbb{E}\left(e^{sX_i}\right)\right) \\
&\le \tfrac{1}{s}\log\left(\textstyle\sum_{i=1}^n e^{\frac{\sigma^2 s^2}{2}}\right) \\
&= \frac{\log n}{s} + \frac{\sigma^2 s}{2}
\end{aligned}$$

- **Theorem 6.** Assume that $\{X_i\}$ are zero-mean random variables, sub-Gaussian with parameter σ. Then

$$\mathbb{E}\left(\max_{i=1,\dots,n} X_i\right) \le \sqrt{2\sigma^2 \log n}, \qquad n \ge 1.$$

- **Proof.** For any $s > 0$,

$$\begin{aligned}
\mathbb{E}\left(\max_{i=1,\dots,n} X_i\right) &= \tfrac{1}{s}\mathbb{E}\left(\log e^{s \max_{1\le i\le n} X_i}\right) \\
&\le \tfrac{1}{s}\log\left(\mathbb{E}\left(e^{s \max_{1\le i\le n} X_i}\right)\right) \\
&= \tfrac{1}{s}\log\left(\mathbb{E}\left(\max\nolimits_{1\le i\le n}\ e^{sX_i}\right)\right) \\
&\le \tfrac{1}{s}\log\left(\mathbb{E}\left(\textstyle\sum_{i=1}^n\ e^{sX_i}\right)\right) \\
&= \tfrac{1}{s}\log\left(\textstyle\sum_{i=1}^n \mathbb{E}\left(e^{sX_i}\right)\right) \\
&\le \tfrac{1}{s}\log\left(\textstyle\sum_{i=1}^n e^{\frac{\sigma^2 s^2}{2}}\right) \\
&= \frac{\log n}{s} + \frac{\sigma^2 s}{2}
\end{aligned}$$

taking $s = \sqrt{2\log n/\sigma^2}$ yields the result.