

# Basic tail and concentration bounds

Kaveh S. Nobari

Lectures in High-Dimensional Statistics

Department of Mathematics and Statistics  
Lancaster University

# Contents

- 1 Classical bounds
  - From Markov to Chernoff
  - Sub-Gaussian variables and Hoeffding bounds
  - Sub-exponential variables and Bernstein bounds
  - Some one-sided results
- 2 Martingale-based methods
- 3 Lipschitz functions of Gaussian variables

# Motivation

It is often of interest to obtain bounds on the tails of a random variable, or two-sided inequalities, which guarantee that the random variable is close to its mean or median. These slides follow the structure of chapter 2 of Wainwright (2019) to shed light on the elementary techniques for obtaining **deviation** and **concentration** inequalities.

One way of controlling a tail probability  $P[X \geq t]$  is by controlling the moments of the random variable  $X$ , where by controlling higher-order moments of the variable  $X$ , we can obtain sharper bounds on tail probabilities. This motivates the "Classical bounds" section of the notes.

We then extend the derivation of bounds to more general functions of the random variables in the "Martingale-based methods" section using **martingale decompositions**, as opposed to limiting the techniques to deriving bounds on **the sum of independent** random variables.

Finally, the seminar is concluded with a classical result on the concentration properties of Lipschitz functions of Gaussian variables.

- 1 Classical bounds
  - From Markov to Chernoff
  - Sub-Gaussian variables and Hoeffding bounds
  - Sub-exponential variables and Bernstein bounds
  - Some one-sided results
- 2 Martingale-based methods
- 3 Lipschitz functions of Gaussian variables

The most elementary tail bound is **Markov's inequality**:

### Markov's inequality

Given a non-negative random variable  $X$  with finite mean - i.e.  $\mathbb{E}[X] < \infty$ , we have

$$P[X \geq t] \leq \frac{\mathbb{E}[X]}{t}, \quad \forall t > 0$$

It is immediately obvious that Markov's inequality **requires only the existence of the first moment**. If the random variable  $X$  also has finite variance - i.e.  $\text{var}(X) < \infty$ , we have **Chebyshev's inequality**:

### Chebyshev's inequality

For a random variable  $X$  that has a finite mean and variance, we have

$$P[|X - \mu| \geq t] \leq \frac{\text{var}(X)}{t^2}, \quad \forall t > 0$$

## Proof.

Chebyshev's inequality follows from Markov's inequality, by considering the variable  $(X - \mathbb{E}[X])^2$  and the constant  $t^2$ . By substituting these in the Markov inequality, we get

$$P[(X - \mathbb{E}[X])^2 \geq t^2] \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2} \quad (1)$$

$$P[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2} \quad (2)$$

Since,  $\mathbb{E}[X] = \mu$  and  $\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$ , we get

$$P[|X - \mu| \geq t] \leq \frac{\text{var}(X)}{t^2}$$



The earlier results can be generalised as follows:

### Extensions of Markov's inequality

Whenever a variable  $X$  has a central moment of order  $k$ , an application of Markov's inequality to the random variable  $|X - \mu|^k$  yields:

$$P[|X - \mu| \geq t] \leq \frac{\mathbb{E}[|X - \mu|^k]}{t^k}, \quad \forall t > 0.$$

and this is not limited to polynomials  $|X - \mu|^k$ :

Suppose  $X$  has a mgf in a neighbourhood of zero, such that there is a constant  $b > 0$  that the functions  $\rho(\lambda) = \mathbb{E}[\exp(\lambda(X - \mu))]$  exists for all  $\lambda < |b|$ . Thus, for any  $\lambda \in [0, b]$ , we may apply Markov's inequality to the random variable  $Y = \exp(\lambda(X - \mu))$ , obtaining the upper bound:

$$P[(X - \mu) \geq t] = P[\exp(\lambda(X - \mu)) \geq \exp(\lambda t)] \leq \frac{\mathbb{E}[\exp(\lambda(X - \mu))]}{\exp(\lambda t)}$$

By taking the log of both side of the latter inequality, we get:

$$\log P[(X - \mu) \geq t] \leq \log \mathbb{E}[\exp(\lambda(X - \mu))] - \lambda t$$

Optimising, our choice of  $\lambda$ , we can obtain the tightest results that yields the Chernoff bound:

### Chernoff bound

$$\log P[(X - \mu) \geq t] \leq \inf_{\lambda \in [0, b]} \{ \log \mathbb{E}[\exp(\lambda(X - \mu))] - \lambda t \}.$$



## 1 Classical bounds

- From Markov to Chernoff
- Sub-Gaussian variables and Hoeffding bounds
- Sub-exponential variables and Bernstein bounds
- Some one-sided results

## 2 Martingale-based methods

## 3 Lipschitz functions of Gaussian variables

Evidently, the form of the tail bound obtained using the Chernoff approach depends on the growth rate of the mgf. Naturally, in the study of the tail bounds the random variables are then classified in terms of their mgfs. The simplest type of behaviour is known as sub-Gaussian, which shall be motivated by deriving tail bounds for a Gaussian variables, say,  $X$ , such that  $X \sim N(\mu, \sigma^2)$ , with density

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(X - \mu)^2}{2\sigma^2}\right)$$

and thus, the mgf

$$\mathbb{E}[\exp(\lambda X)] = \exp\left(\mu\lambda + \frac{\sigma^2\lambda^2}{2}\right), \quad \forall \lambda \in \mathbb{R}$$

## Example (Gaussian tail bounds)

Let  $X \sim N(\mu, \sigma^2)$  be a Gaussian r.v., which has mgf

$$\mathbb{E}[\exp(\lambda X)] = \exp\left(\mu\lambda + \frac{\sigma^2\lambda^2}{2}\right), \quad \forall \lambda \in \mathbb{R}$$

substituting this into the optimising problem of the Chernoff bound, we get

$$\inf_{\lambda \geq 0} \{\log \mathbb{E}[\exp(\lambda(X - \mu))] - \lambda t\} = \inf_{\lambda \geq 0} \left\{ \frac{\sigma^2\lambda^2}{2} - \lambda t \right\} = -\frac{t^2}{2\sigma^2} \quad (3)$$

Therefore, we can conclude that any  $N(\mu, \sigma^2)$  r.v. satisfies the upper deviation inequality

$$P[X \geq \mu + t] \leq \exp\left(-\frac{t^2}{2\sigma^2}\right) \quad (4)$$

## Proof of equation (3).

To solve the optimisation problem below

$$\inf_{\lambda \geq 0} \left\{ \frac{\sigma^2 \lambda^2}{2} - \lambda t \right\}$$

we take derivatives to find the optimum of this quadratic function, - i.e.

$$\frac{\partial}{\partial \lambda} \left( \frac{\sigma^2 \lambda^2}{2} - \lambda t \right) = 0,$$

which leads to  $\lambda_{opt} = \frac{t}{\sigma^2}$ . Substituting  $\lambda_{opt}$  with  $\lambda$  in the above equation yields relationship (3). □

## Definition (Sub-Gaussianity)

A r.v.  $X$  with mean  $\mu = \mathbb{E}[X]$  is sub-Gaussian if there is a positive number  $\sigma$ , such that

$$\mathbb{E}[\exp(\lambda(X - \mu))] \leq \exp\left(\frac{\sigma^2 \lambda^2}{2}\right), \quad \forall \lambda \in \mathbb{R}$$

where the constant  $\sigma$  is referred to as the **sub-Gaussian parameter**. Moreover, by the symmetry of the definition, the variable  $-X$  is sub-Gaussian iff  $X$  is sub-Gaussian, so that we also have lower deviation inequality  $P[X \leq \mu - t] \leq \exp\left(-\frac{t^2}{2\sigma^2}\right)$ ,  $\forall t \geq 0$ . Thus, we conclude that any sub-Gaussian variable satisfies the **concentration inequality**

$$P[|X - \mu| \geq t] \leq 2 \exp\left(-\frac{t^2}{2\sigma^2}\right), \quad \forall t \in \mathbb{R}.$$

We may have scenarios in which sub-Gaussian variables are non-Gaussian.

### Example (Rademacher variables)

A Rademacher r.v.  $\varepsilon$  takes the values  $[-1, +1]$  equiprobably -i.e  $P[\varepsilon = -1] = P[\varepsilon = +1] = \frac{1}{2}$ . Thus, the mgf of  $\varepsilon$  is as follows

$$\mathbb{E}[\exp(\lambda\varepsilon)] = \sum_{i \in \{-1, +1\}} \exp(\lambda\varepsilon_i) p(\varepsilon = i) = \frac{1}{2} [\exp(-\lambda) + \exp(\lambda)]$$

where the Maclaurin-series expansion of the terms  $\exp(-\lambda)$  and  $\exp(\lambda)$  leads gives us

$$\begin{aligned} \mathbb{E}[\exp(\lambda\varepsilon)] &= \frac{1}{2} \left[ \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} + \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} \right] = \frac{1}{2} \left[ 2 \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{2k!} \right] = \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{2k!} \\ &\leq 1 + \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{2^k k!} = \exp\left(\frac{\lambda^2}{2}\right) \end{aligned}$$

with the sub-Gaussian parameter  $\sigma = 1$ .

**Some preliminaries:** The exponential function  $g(z) = \exp(z)$  is convex; thus, Jensen's inequality for convex functions applies as follows

$$g(\mathbb{E}[z]) \leq \mathbb{E}[g(z)]$$

A r.v.  $Z'$  is an **independent copy** of  $Z$ , if it has the same distribution as  $Z$ , and where  $Z$  and  $Z'$  are independent.

Given the above definitions, we provide a simple example of **symmetrization argument**, in which first an independent copy of  $X$ ,  $X'$  is introduced and the problem is symmetrized using a Rademacher variable.

## Symmetrization argument

Let  $X$  be a r.v. with mean zero - i.e.  $\mu = \mathbb{E}_X[X] = 0$ , with a support on the interval  $[a, b]$ , and let  $X'$  be an independent copy of  $X$ , for any  $\lambda \in \mathbb{R}$ , we have

$$\mathbb{E}_X[\exp(\lambda X)] = \mathbb{E}_X[\exp(\lambda(X - \mathbb{E}_{X'}[X']))]$$

since  $\mathbb{E}_X[X] = \mathbb{E}_{X'}[X'] = 0$ . Using Jensen's inequality, we further establish that

$$\mathbb{E}_X[\exp(\lambda(X - \mathbb{E}[X']))] \leq \mathbb{E}_{X, X'}[\exp(\lambda(X - X'))]$$

Further, note that  $\varepsilon(X - X')$  and  $(X - X')$  possess the same distribution, where  $\varepsilon$  is a Rademacher r.v., so that

$$\mathbb{E}_{X, X'}[\exp(\lambda(X - X'))] = \mathbb{E}_{X, X'}[\mathbb{E}_\varepsilon[\exp(\lambda\varepsilon(X - X'))]],$$



where from the earlier example, we know that

$$\mathbb{E}_{X, X'} [\mathbb{E}_{\varepsilon} [\exp(\lambda \varepsilon (X - X'))]] \leq \mathbb{E}_{X, X'} \left[ \exp \left( \frac{\lambda^2 (X - X')^2}{2} \right) \right]$$

since  $|X - X'| \leq b - a$ , we are guaranteed that

$$\mathbb{E}_{X, X'} \left[ \exp \left( \frac{\lambda^2 (X - X')^2}{2} \right) \right] \leq \exp \left( \frac{\lambda^2 (b - a)^2}{2} \right)$$

thus, we have shown that  $X$  is sub-Gaussian with sub-Gaussian parameter  $\sigma = b - a$

## Quiz:

- 1) Two independent sub-Gaussian variables  $X_1$  and  $X_2$  possess the sub-Gaussian parameters  $\sigma_1$  and  $\sigma_2$  respectively. What is the sub-Gaussian parameter of  $X_1 + X_2$ ?
- 2) Now once again consider the sub-Gaussian tail bound (4). How is this result extended to the variable  $X_1 + X_2$ ?

The answers to the above quiz, can be generalised to the variables  $X_1, \dots, X_n$  with mean  $\mu_i$  and sub-Gaussian parameters  $\sigma_i$  for  $i = 1, \dots, n$  leading to the Hoeffding bound

### Hoeffding bounds

Suppose that the variables  $X_1, \dots, X_n$  each with mean  $\mu_1, \dots, \mu_n$  and sub-Gaussian parameter  $\sigma_1, \dots, \sigma_n$  are independent. Then we have

$$P \left[ \sum_{i=1}^n (X_i - \mu_i) \geq t \right] \leq \exp \left\{ -\frac{t^2}{2 \sum_{i=1}^n \sigma_i^2} \right\}$$

To prove the equivalent characterizations of sub-Gaussian variables, it is of interest to first answer Exercise 2.2 of Wainwright (2019) which introduces **Mills ratio**.

**Exercise 2.2 of Wainwright (2019):** Let  $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right)$  be the density function of a standard normal  $Z \sim N(0, 1)$  variate.

- 1) Show that  $\phi'(z) + z\phi(z) = 0$
- 2) Use part 1 to show that

$$\phi(z) \left( \frac{1}{z} - \frac{1}{z^3} \right) \leq P[Z \geq z] \leq \phi(z) \left( \frac{1}{z} - \frac{1}{z^3} + \frac{3}{z^5} \right), \quad \forall z > 0$$

**Solution:**

Part 1:

$$\phi'(z) = -\frac{z}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \quad \text{and} \quad z\phi(z) = \frac{z}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right)$$

thus,

$$\phi'(z) + z\phi(z) = 0$$

## Part 2:

Note that  $P[Z \geq z] = \int_z^\infty \phi(t) dt$ . Furthermore, from part 1, we know that  $\phi(z) = \frac{-\phi'(z)}{z}$ . By substituting  $\frac{-\phi'(z)}{z}$  into the earlier integral, we get

$$\int_z^\infty \phi(t) dt = \int_z^\infty \frac{-\phi'(t)}{t} dt = \left[ \frac{-\phi'(t)}{t} \right]_z^\infty - \int_z^\infty \frac{\phi(t)}{t^2} dt$$

We know that  $\lim_{t \rightarrow \infty} \frac{-\phi'(t)}{t} = 0$ , therefore, we may apply the the above expression can be written as

$$\frac{\phi(z)}{z} - \int_z^\infty \frac{-\phi'(t)}{t^3} dt$$

using the substitution derived from Miller's ratio. Using integration by parts yet again, we obtain

$$\begin{aligned}\frac{\phi(z)}{z} - \int_z^\infty \frac{-\phi'(t)}{t^3} dt &= \frac{\phi(z)}{z} + \left[ \frac{\phi(t)}{t^3} \right]_z^\infty - \int_z^\infty \frac{-3\phi(t)}{t^4} dt \\ &= \frac{\phi(z)}{z} + \frac{\phi(z)}{z^3} + \int_z^\infty \frac{3\phi(t)}{t^4} dt \\ P[Z \geq z] &= \phi(z) \left( \frac{1}{z} + \frac{1}{z^3} \right) + \underbrace{\int_z^\infty \frac{3\phi(t)}{t^4} dt}_{\geq 0} \\ &\geq \phi(z) \left( \frac{1}{z} + \frac{1}{z^3} \right)\end{aligned}$$

Applying the same procedure again will prove the upper inequality. This is left as an exercise to the reader.

## Equivalent characterizations of the sub-Gaussian variables (I-II)

- (I) From the definition of sub-Gaussian variables, a r.v. with  $\mu = \mathbb{E}[X] = 0$  is sub-Gaussian for  $\sigma \geq 0$ ,

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

- (II) There is a constant  $c \geq 0$  and Gaussian r.v.  $Z \sim N(0, \tau^2)$ , such that

$$P[|X| \geq s] \leq cP[|Z| \geq s], \quad \forall s \geq 0.$$

- 1 Classical bounds
  - From Markov to Chernoff
  - Sub-Gaussian variables and Hoeffding bounds
  - Sub-exponential variables and Bernstein bounds
  - Some one-sided results
- 2 Martingale-based methods
- 3 Lipschitz functions of Gaussian variables



The notion of sub-Gaussianity is rather restrictive. We thus now introduce sub-exponential variables, which impose milder conditions on the mgf.

### Definition

**Sub-exponentiality** A r.v.  $X$  with mean  $\mu = \mathbb{E}[X]$  is sub-exponential if there are non-negative parameters  $(\nu, \alpha)$ , such that

$$\mathbb{E}[\exp(\lambda(X - \mu))] \leq \exp\left(\frac{\nu^2 \lambda^2}{2}\right), \quad \forall |\lambda| < \frac{1}{\alpha}$$

It is immediately obvious that any sub-Gaussian variable is also sub-exponential, where the former is a special case of the latter, with  $\nu = \sigma$  and  $\alpha = 0$ . However, the converse is not true.

An example of a case where a variable is sub-exponential but not sub-Gaussian is as follows

## Example (sub-exponential but not sub-Gaussian)

Let  $Z \sim N(0, 1)$ , and consider the r.v.  $X = Z^2$ , such that  $Z \sim \chi_1^2$ . Therefore, the mean  $\mu = \mathbb{E}[X] = 1$ . For  $\lambda < \frac{1}{2}$ , we have the mgf as follows

$$\begin{aligned}\mathbb{E}[\exp(\lambda(X - 1))] &= \int_{-\infty}^{+\infty} \exp(\lambda(Z^2 - 1)) f(z) dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(\lambda(Z^2 - 1)) \exp\left(-\frac{Z^2}{2}\right) dz \\ &= \frac{\exp(-\lambda)}{\sqrt{1 - 2\lambda}}.\end{aligned}$$

for  $\lambda \geq \frac{1}{2}$  the mgf is infinite, which reveals that  $X$  is not sub-Gaussian.

To obtain the tail-bounds of sub-exponential variables, we refer to the Chernoff-type approach - i.e.

$$P[X - \mu \geq t] = P[\exp(\lambda(X - \mu)) \geq \exp(t\lambda)] \leq \frac{\mathbb{E}[\exp(\lambda(X - \mu))]}{\exp(\lambda t)}$$

where from the definition of sub-exponential variables, we get the upper bound

$$P[X - \mu \geq t] \leq \frac{\mathbb{E}[\exp(\lambda(X - \mu))]}{\exp(\lambda t)} \leq \exp\left(\frac{\lambda^2 \nu^2}{2} - \lambda t\right), \quad \forall \lambda \in \left[0, \frac{1}{\alpha}\right],$$

where the Chernoff optimisation problem is

$$\log P[X - \mu \geq t] \leq \inf_{\lambda \in [0, \alpha^{-1}]} \left\{ \frac{\lambda^2 \nu^2}{2} - \lambda t \right\}$$

where using the same unconstrained optimisation approach as for sub-Gaussian variables, we'd obtain  $\lambda_{opt} = \frac{t}{\nu^2}$ , which yields the minimum  $-\frac{t^2}{2\nu^2}$ .

Recall the constraint  $0 \leq \lambda < \frac{1}{\alpha}$ . This implies that the unconstrained optimal  $\lambda_{opt}$  must be between  $0 \leq \frac{t}{\nu^2} < \frac{1}{\alpha}$ , which implies that in the interval  $0 \leq t < \frac{\nu^2}{\alpha^2}$ , the unconstrained optimum corresponds to the constrained optimum.

Otherwise for  $t \geq \frac{\nu^2}{\alpha^2}$ , considering that the function  $g(\cdot, t) = \frac{\lambda^2 \nu^2}{2} - \lambda t$  is monotonically decreasing, in the interval  $[0, \lambda^*)$ , the constrained minimum is obtained at the boundary - i.e.  $\lambda^\# = \frac{1}{\alpha}$ , which leads to the minimum

$$g^*(t) = g(\lambda^\#, t) = -\frac{t}{\alpha} + \frac{1}{2\alpha} \frac{\nu^2}{\alpha} \leq -\frac{t}{2\alpha}$$

where this inequality used the fact that  $\frac{\nu^2}{\alpha} \leq t$ .

The results above lead to the sub-exponential tail bounds as follows

## Sub-exponential tail bounds

Suppose  $X$  is sub-exponential with parameters  $(\nu, \alpha)$ . Then

$$P[X - \mu \geq t] \leq \begin{cases} \exp\left(-\frac{t}{2\nu^2}\right) & 0 \leq t \leq \frac{\nu^2}{\alpha}, \\ \exp\left(-\frac{t}{2\alpha}\right) & t > \frac{\nu^2}{\alpha}. \end{cases}$$

The sub-exponential property can be verified by computing or bounding the mgf, which may not be practical in many different settings. One other approach is based on the control of the polynomial moments of  $X$ , which leads to the **Bernstein condition**

## Bernstein condition

Given a r.v.  $X$  with mean  $\mu = \mathbb{E}[X]$  and variance  $\sigma^2 = \mathbb{E}[X^2] - \mu^2$ , the Bernstein condition with parameter  $b$  holds if

$$|\mathbb{E}[(X - \mu)^k]| \leq \frac{1}{2} k! \sigma^2 b^{k-2}, \quad k \geq 2$$

One sufficient condition for the Bernstein condition to hold is that  $X$  is bounded. When  $X$  satisfies the Bernstein condition, then it is sub-exponential with parameters  $\sigma^2$  and  $b$ . The Maclaurin-series expansion of the mgf can be expressed as follows

$$\begin{aligned}\mathbb{E}[\exp(\lambda(X - \mu))] &= \mathbb{E}\left\{\sum_{i=0}^{\infty} \frac{f^{(i)}(0)}{i!} \lambda(X - \mu)^i\right\} = \sum_{i=0}^{\infty} \mathbb{E}\left\{\frac{f^{(i)}(0)}{i!} \lambda(X - \mu)^i\right\} \\ &= 1 + \lambda \mathbb{E}[(X - \mu)] + \frac{\lambda^2 \mathbb{E}[(X - \mu)]^2}{2} + \sum_{i=3}^{\infty} \frac{\lambda^i \mathbb{E}[(X - \mu)^i]}{i!} \\ &= 1 + \frac{\lambda^2 \sigma^2}{2} + \sum_{i=3}^{\infty} \frac{\lambda^i \mathbb{E}[(X - \mu)^i]}{i!}\end{aligned}$$

## 1 Classical bounds

- From Markov to Chernoff
- Sub-Gaussian variables and Hoeffding bounds
- Sub-exponential variables and Bernstein bounds
- Some one-sided results

## 2 Martingale-based methods

## 3 Lipschitz functions of Gaussian variables

- 1 Classical bounds
  - From Markov to Chernoff
  - Sub-Gaussian variables and Hoeffding bounds
  - Sub-exponential variables and Bernstein bounds
  - Some one-sided results
- 2 Martingale-based methods
- 3 Lipschitz functions of Gaussian variables



- 1 Classical bounds
  - From Markov to Chernoff
  - Sub-Gaussian variables and Hoeffding bounds
  - Sub-exponential variables and Bernstein bounds
  - Some one-sided results
- 2 Martingale-based methods
- 3 Lipschitz functions of Gaussian variables

# References

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press.