High-Dimensional Statistics: A Non-Asymptotic Viewpoint

Chapter 2: Basic tail and concentration bounds

Scribe: Renxiong Liu

For high-dimensional statistics, we will often be confronted with theorems on "bounds". In this chapter, we explore some elementary techniques for obtaining both deviation and concentration inequalities. These techniques will serve as the starting points for large-deviation bounds and concentration of measure.

2.1 Classical bounds

The most elementary bound that we have seen is given by the *Markov inequality*, which often provides a relative loose result in terms of bounding.

Theorem 2.1 (Markov inequality). *Let X be a non-negative random variable. For any $t > 0$, we have that*

$$\mathbb{P}(X > t) \leq \frac{\mathbb{E}X}{t} \tag{2.1}$$

Several natural corollary of *Markov inequality* (c.f. Theorem 2.1) is given as below.

Corollary 2.1. *Let X be a random variable such that $\mathbb{E}|X|^k < \infty$, then for all $t > 0$*

1. *(Chebyshevs inequality) Suppose $k \geq 2$. Let $\mu = \mathbb{E}X$, then*

$$\mathbb{P}(|X - \mu| > t) \leq \frac{var(X)}{t^2} \tag{2.2}$$

2. *Suppose $k \geq 1$, Let $\mu = \mathbb{E}X$, then for all $t > 0$*

$$\mathbb{P}(|X - \mu| > t) \leq \frac{\mathbb{E}|X - \mu|^k}{t^k}$$

The other classic bound is called *Chernoff bound*, which is a generic method to construct the bound.

Theorem 2.2 (Chernoff-bound). *Let X be a random variable such that $\mathbb{E}|X| < \infty$. Then for any $t \in \mathbb{R}$, we have that*

$$\mathbb{P}(X - \mathbb{E}X > t) \leq \min_{\lambda > 0} \frac{1}{e^{\lambda t}} \mathbb{E}e^{\lambda(X - \mathbb{E}X)} \tag{2.3}$$

In next section, we will see that Chernoff bounding trick serves as the most basic technique to control the tail probability.

2.2 Sub-Gaussian variables and Hoeffding bound

In this section, we introduce sub-Gaussian random variables and present one widely discussed bound, named *Hoeffding bound*.

Definition 2.1 (sub-Gaussian variables)**.** A random variable X is sub-Gaussian with parameter σ if for any $\lambda \in \mathbb{R}$,

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

Note for any Gaussian random variable $X \in N(\mu, \sigma^2)$, X must be a sub-Gaussian with parameter σ. Hence the definition of sub-Gaussian variables extends the scope of discussed random variables from Gaussian to a larger class. Some examples are given below.

Example 2.1 (Rademacher variables)**.** A Rademacher random variable ϵ takes value $\{-1, 1\}$ with equal probability. In this example, we show that ϵ is sub-Gaussian variable with $\sigma = 1$.

Proof. We use series expansion to show this claim. We have that

$$\mathbb{E}e^{\lambda(\epsilon-\mathbb{E}\epsilon)} = \mathbb{E}e^{\lambda\epsilon} = \mathbb{E}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\epsilon^k = \mathbb{E}\sum_{k=0}^{\infty}\frac{\lambda^{2k}}{(2k)!}\epsilon^{2k} = \sum_{k=0}^{\infty}\frac{\lambda^{2k}}{(2k)!} \le \sum_{k=0}^{\infty}\frac{\lambda^{2k}}{2^k k!} = e^{\frac{1}{2}\lambda^2}$$

This completes the proof.

Example 2.2 (bounded variables)**.** To simplify the proof, let X be a random variable such that $X \in [a, b]$ and $\mathbb{E}X = 0$. We show that X is sub-Gausian with $\sigma = (b - a)$.

Proof. We apply the *symmetrization argument* to show this claim. Let X' be an independent copy of X, that is, X, X' are i.i.d. We have that

$$\mathbb{E}e^{\lambda(X-\mathbb{E}X)} = \mathbb{E}e^{\lambda(X-\mathbb{E}X')} = \mathbb{E}e^{\lambda\mathbb{E}[X-X'|X]} \le \mathbb{E}e^{\lambda(X-X')}$$

Here we apply *Jensen's inequality*. To proceed, let ϵ be a Rademacher variable and independent with X, X'. Note

$$\mathbb{E}e^{\lambda\epsilon(X-X')} = \mathbb{E}\{\mathbb{E}[e^{\lambda\epsilon(X-X')} \mid \epsilon]\} = \frac{1}{2}\mathbb{E}e^{\lambda\epsilon(X-X')} + \frac{1}{2}\mathbb{E}e^{\lambda\epsilon(X'-X)} = \mathbb{E}e^{\lambda(X-X')}$$

We have that

$$\mathbb{E}e^{\lambda(X-X')} = \mathbb{E}\{\mathbb{E}[e^{\lambda\epsilon(X-X')} \mid X, X']\} \le \mathbb{E}\{e^{\frac{1}{2}\lambda^2(X-X')^2}\} \le e^{\frac{1}{2}(b-a)^2\lambda^2}$$

This completes the proof.

Next, we present the useful Hoeffding bound and give some remarks on how to use that.

Theorem 2.3 (Hoeffding bound)**.** *Let* $\{X_i\}_{i=1}^n$ *be independent sub-Gaussian random variables with parameters* $\{\sigma_i\}_{i=1}^n$*. Then for any* $t > 0$*, we have that*

$$\mathbb{P}(\sum_{k=1}^{n}(X_k - \mathbb{E}X_k) > t) \le \exp(-\frac{t^2}{2\sum_{k=1}^n \sigma_k^2}) \tag{2.4}$$

Proof. By Chernoff-bound trick, for $t > 0$, we have that

$$\mathbb{P}(\sum_{k=1}^{n}(X_k - \mathbb{E}X_k) > t) \leq \frac{1}{\exp(\lambda t)} \prod_{k=1}^{n} \mathbb{E}e^{\lambda(X_k - \mathbb{E}X_k)} \leq \exp(\frac{1}{2}\lambda^2 \sum_{k=1}^{n} \sigma_k^2 - \lambda t)$$

Note

$$\min_{\lambda > 0} \exp(\frac{1}{2}\lambda^2 \sum_{k=1}^{n} \sigma_k^2 - \lambda t) = \exp(-\frac{t^2}{2\sum_{k=1}^{n} \sigma_k^2})$$

This finishes the proof.

Remark. We leave out Theorem 2.6 in the book, which states some equivalent form of defining sub-Gaussian.

2.3 Sub-exponential variables and Bernstein bounds

In this section, we extend the scope of discussed random variables from sub-Gaussian to sub-exponential and present the Bernstain bound, which is widely used in literatures.

Definition 2.2 (sub-exponential variables)**.** A random variable X is sub-exponential with parameter (σ, α) if for any $|\lambda| \leq \frac{1}{\alpha}$,

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

It's easy to see that if X is a sub-Gaussian variable with parameter σ, then it's a sub-exponential variable with parameters $(\sigma, 0)$. In next example, we will show that a sub-exponential variable is not necessary a sub-Gaussian variable, which means the family of sub-exponential variables is larger than the family of sub-Gaussian variables.

Example 2.3 (sub-exponential is not sub-Gaussian)**.** Let $Z \sim N(0, 1)$ be a standard normal random variable. Denote $X = Z^2$, in this example, we show

(1). X is not sub-Gaussian

(2). X is sub-exponential with parameter $(\sigma, \alpha) = (2, 4)$.

Proof. To see this result, note $\mathbb{E}X = 1$, we have that

$$\mathbb{E}\exp(\lambda(X-1)) = \int_{\mathbb{R}} e^{\lambda(z^2-1)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \leq \exp(\frac{1}{2} \cdot 4\lambda^2), \text{ for } |\lambda| < \frac{1}{4}$$

which validates (2). To see why X is not sub-Gaussian, note

$$\mathbb{E}\exp(\lambda(X-1)) = \frac{e^{-\lambda}}{\sqrt{1-2\lambda}}$$

when $\lambda \to 1/2^+$, we will have $\mathbb{E}\exp(\lambda(X-1)) \to \infty$, which is a contradiction.