Chapter 14
Basic concentration Inequalities

We are now in the position to prove a concentration result for a sum of sub-gaussian random variables.

Theorem 14.1 (General Hoeffding inequality). *Assume $m_1, \dots, m_N$ are independent, sub-gaussian random variables, then there exists a constant c such that the following concentration inequality holds:*

$$\mathbb{P}\left[\left|\sum_{i=1}^{N} \mathbf{a}_i m_i\right| > t\right] \leq 2e^{-\frac{t^2}{c^2\|\mathbf{a}\|^2}}, \quad \forall t \geq 0.$$

Proof. Using the Chernoff inequality we have

$$\mathbb{P}\left[\sum_{i=1}^{N} \mathbf{a}_i m_i \geq t\right] \leq \min_{\lambda>0} \frac{\mathbb{E}\left[e^{\lambda \sum_{i=1}^{N} \mathbf{a}_i m_i}\right]}{e^{\lambda t}} \leq \min_{\lambda>0} \frac{e^{\lambda^2 \sum_{i=1}^{N} c_i^2 \mathbf{a}_i^2}}{e^{\lambda t}} = e^{-\frac{t^2}{2\sum_{i=1}^{N} c_i^2 \mathbf{a}_i^2}},$$

where we have used Proposition 13.2 in the third inequality. Similarly, we can show that

$$\mathbb{P}\left[\sum_{i=1}^{N} \mathbf{a}_i m_i \leq -t\right] \leq e^{-\frac{t^2}{2\sum_{i=1}^{N} c_i^2 \mathbf{a}_i^2}},$$

which together with the previous inequality concludes the proof.

Corollary 14.1. *Show that there exists a constant c such that*

$$\mathbb{P}\left[\left|\frac{1}{N}\sum_{i=1}^{N} m_i\right| > t\right] \leq 2e^{-N\frac{t^2}{c^2}}, \quad \forall t \geq 0.$$

The results of Corollary 14.1 is a direct extension of concentration phenomenon observed from Gaussian random variables that we discuss at the beginning of this section. This, together with the LLN, says that the sample mean not only converges almost surely to the mean but also concentrates at the mean. The Hoeffding inequality quantifies this concentration via the exponential decaying tail probability.

Exercise 14.1. Let m be a Rademacher (aka symmetric Bernoulli) random variables , i.e., probability of m being either 1 or -1 is 1/2. Show that *Rademacher distribution is also a sub-gaussian*, i.e.,

$$\mathbb{E}\left[e^{\lambda m}\right] \leq e^{\frac{\lambda^2}{2}}. \tag{14.1}$$

Exercise 14.2 (Hoeffding Lemma). Let m be a random variable such that $\mathbb{E}[m] = 0$ and $m \in [a,b]$ almost surely. Show that

$$\mathbb{E}\left[e^{\lambda m}\right] \leq e^{\lambda^2 \frac{(b-a)^2}{8}}.$$

In particular, *mean zero bounded random variables are subgaussian*.

Hint. With the help of Jensen inequality, show that

$$\mathbb{E}_m\left[e^{\lambda m}\right] = \mathbb{E}_m\left[e^{\lambda \mathbb{E}_{m'}[m-m']}\right] \leq \mathbb{E}_{m,m'}\left[e^{\lambda\left(m-m'\right)}\right],$$

and then show that

$$\mathbb{E}_{m,m'}\left[e^{\lambda\left(m-m'\right)}\right] = \mathbb{E}_{m,m'}\left[\mathbb{E}_y\left[e^{\lambda y\left(m-m'\right)}\right]\right],$$

where y is a Rademacher random variable. Then finish the work with the help of (14.1).

Exercise 14.3 (Hoeffding inequality for sum of bounded random variables). Let $m_i, i = 1,\ldots,N$ be independent bounded random variables such that $m_i \in [a_i, b_i]$ almost surely. Show that

$$\mathbb{P}\left[|S_N - \mathbb{E}[S_N]| > t\right] \leq 2e^{-2N^2 \frac{t^2}{\sum_{i=1}^N (b_i-a_i)^2}}, \quad \forall t \geq 0,$$

where $S_N = \frac{1}{N}\sum_{i=1}^N m_i$.

Deduce a concentration inequality for $\sum_{i=1}^N m_i$.

Application. Toss a fair coin N time, what is the probability of getting at least $3N/4$ heads?

Definition 14.1 (ℓ-percent sparse random variables [20, 18]).

Let $s = \frac{1}{1-\ell}$ where $\ell \in [0,1)$ is the level of sparsity desired. Then

$$\zeta = \sqrt{s}\begin{cases} +1 & \text{with probability } \frac{1}{2s}, \\ 0 & \text{with probability } \ell = 1 - \frac{1}{s}, \\ -1 & \text{with probability } \frac{1}{2s} \end{cases} \tag{14.2}$$

is a *ℓ-percent sparse distribution.*

Note that for $\ell = 0$, ζ corresponds to a Rademacher distribution, and that $\ell = 2/3$ corresponds to the *Achlioptas distribution* [1]. By inspection we have that $\mathbb{E}[\zeta] = 0$ and $\mathbb{E}\left[\zeta^2\right] = 1$. As shall be shown, this class of random variable is important for large-scale application involving random vectors/matrices, since vectors/matrices generated from this class of randomn variables can be very sparse depending on ℓ.

Exercise 14.4. Show that $\mathbb{E}[\zeta] = 0$ and $\mathbb{E}\left[\zeta^2\right] = 1$. Morever, show that *$\ell$-percent sparse random variables are sub-gaussian* and determine their proxy.

In practice, random variables do not always have zero mean. In that case we can center a random variable by substracting out its mean. We can then define a random variable m with mean $\overline{m}$ as a sub-gaussian with proxy σ^2 by saying that $m - \overline{m}$ is a sub-gaussian, i.e.,

$$\mathbb{E}\left[e^{\lambda(m-\overline{m})}\right] \leq e^{\frac{\lambda^2\sigma^2}{2}}.$$

All the above results are valid for $m - \overline{m}$.

Exercise 14.5. Prove a concentration result similar to that of Corollary 14.1 for the case when $m_i - \overline{m}_i$ are sub-gaussian with proxy σ_i^2.