Chapter 12
Concentration of Gaussian random variables

12.1 Important mathematical preliminaries

Lemma 12.1 (Markov inequality). *Let m be a non-negative random variable. There holds:*

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

Proof. We have

$$\mathbb{E}[m] = \mathbb{E}\left[m\mathbb{1}_{\{m\geq t\}}\right] + \mathbb{E}\left[m\mathbb{1}_{\{m<t\}}\right] \geq t\mathbb{E}\left[\mathbb{1}_{\{m\geq t\}}\right] \stackrel{\text{def}}{=} t\mathbb{P}[m \geq t],$$

where

$$\mathbb{1}_{\{m\geq t\}} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } m \geq t \\ 0 & \text{otherwise} \end{cases}.$$

Exercise 12.1 (Chebyshev and Chernoff inequalities). Using Markov inequality to show the following Chebyshev inequality:

$$\mathbb{P}[|m| \geq t] \leq \frac{\sigma^2}{t^2}, \quad \forall t > 0,$$

for any zero mean random variable m with variance σ^2.

Can you derive a Chebyshev inequality when $\mathbb{E}[m] = \overline{m}$?

We can see that the Chebyshev inequality improves the Markov inequality using the second moment. In fact we can use all the moments to drastically sharpen the inquality, and this due to Chernoff.

Show that, by using the monoticity of the exponential funciton (instead of the square function), there holds

$$\mathbb{P}[m \geq t] \leq \min_{\lambda>0} \frac{\mathbb{E}\left[e^{\lambda m}\right]}{e^{\lambda t}}, \quad \forall t > 0, \tag{12.1}$$

for any non-negative random variable m. That is, assume the *moment generating function (MGF)* $\mathbb{E}\left[e^{\lambda m}\right]$ is bounded, the tail probability of m decays exponentially.

This “concentration phenomenon” will be made precise in this chapter for a large class of sub-gaussian random variables with bounded MGF. Exercise 12.2 provides a concentration result for Gaussian variables.

Exercise 12.2 (A tail bound of normal distribution). We now apply the Chernoff inequality (12.1) to normal random variables, i.e. $m \sim \mathcal{N}\left(\overline{m}, \sigma^2\right)$.

1. Show that the MGF of m is given by

$$\mathbb{E}\left[e^{\lambda m}\right] = e^{\lambda \overline{m} + \frac{\lambda^2 \sigma^2}{2}}, \quad \forall t \in \mathbb{R}.$$

2. Solve the optimization on the right hand side of (12.1) to conclude that

$$\mathbb{P}\left[m - \overline{m} \geq t\right] \leq e^{-\frac{t^2}{2\sigma^2}}, \quad \forall t \geq 0. \tag{12.2}$$

Proposition 12.1 (Connection between tail bound and expectation). *For any random variable m, there holds:*

$$\mathbb{E}\left[m\right] = \int_0^\infty \mathbb{P}\left[m > t\right] dt - \int_{-\infty}^0 \mathbb{P}\left[m < t\right] dt.$$

Proof. Assume that $m \geq 0$, we have

$$m = \int_0^m dt = \int_0^\infty \mathbb{1}_{\{m > t\}} dt.$$

Taking expectation both sides and with the help with Fubini theorem we conclude

$$\mathbb{E}\left[m\right] = \int_0^\infty \mathbb{P}\left[m > t\right] dt.$$

Do you see this?

The proof for the general case is similar by noting that

$$m = m\mathbb{1}_{m \geq 0} + m\mathbb{1}_{m < 0}.$$

Lemma 12.2 (Hutchinson 1989). *Let* $\mathbf{m}$ *be an n-dimensional random vector with* $\mathbb{E}\left[\mathbf{m}\right] = \mathbf{0}$ *and* $\mathbb{V}ar\left[\mathbf{m}\right] = \mathbb{E}\left[\mathbf{m}\mathbf{m}^T\right] = \mathbf{I}$. *Then:*

$$Tr\left(\mathscr{A}\right) = \mathbb{E}\left[\mathbf{m}^T \mathscr{A} \mathbf{m}\right],$$

for any matrix $\mathscr{A} \in \mathbb{R}^{n \times n}$.

Do you see this?

Proof. We have

$$Tr\left(\mathscr{A}\right) = Tr\left(\mathscr{A}\mathbb{V}ar\left[\mathbf{m}\right]\right) = \mathbb{E}\left[Tr\left(\mathscr{A}\mathbf{m}\mathbf{m}^T\right)\right] = \mathbb{E}\left[Tr\left(\mathbf{m}^T \mathscr{A} \mathbf{m}\right)\right] = \mathbb{E}\left[\mathbf{m}^T \mathscr{A} \mathbf{m}\right].$$

A useful bound that we use repeatedly is the *union bound*.

Lemma 12.3 (Union bound).

$$\mathbb{P}[A_1 \cup A_2 \cup \ldots \cup A_n] \leq \sum_{i=1}^{n} \mathbb{P}[A_i], \quad \forall n \in \mathbb{N}.$$

12.2 Concentration of sum of scalar Gaussian random variables

Let us start with an interesting application of the Chebyshev inequality, namely, a version of the weak law of large numbers. Let $m_i \sim \mathcal{N}\left(\overline{m}, \sigma^2\right)$, and $S_N \stackrel{\text{def}}{=} \frac{1}{N}(m_1 + m_2 + \cdots + m_N)$. We can show

Can you easily show this?

$$S_N \sim \mathcal{N}\left(\overline{m}, \sigma^2/N\right),$$

which together with the Chebyshev inequality leads to

$$\mathbb{P}[|S_N - \overline{m}| \geq t] \leq \frac{1}{N}\frac{\sigma^2}{t^2},$$

which in turns yields

$$\lim_{N\to\infty} \mathbb{P}[|S_N - \overline{m}| \geq t] = 0, \quad \forall t > 0,$$

that is, the probability that S_N deviates from its mean approaches 0 as the number of i.i.d random summands increases. This is of course not surprising since we already know from the LLN that S_N indeed converges to $\overline{m}$ almost surely. What is more interesting is whether the quadratic decay of the deviation is sharp. The answer is that it is very conservative in this case. To see this, we apply (12.2) to $S_N - \overline{m}$ to obtain

$$\mathbb{P}[|S_N - \overline{m}| \geq t] \leq 2e^{-N\frac{t^2}{2\sigma^2}},$$

which shows that *a t-deviation of S_N from its mean decays exponentially in both t and N*.

In this chapter by concentration we refer to the phenomenon that a random variable concentrates around its mean. We quantify this phenomenon via concentration inequality of the form

$$\mathbb{P}[|m - \overline{m}| \geq t] \leq \text{ small quantity}, \tag{12.3}$$

i.e. the *tail bound*.

Remark 12.1. It is important to see that while results from CLT and LLN are asymptotic (valid for $N \to \infty$), concentration inequalities are non-asymptotic (valid for finite N).

We have shown that normal random variables have exponential tail bounds which decay rapidly (the best possible we argue). It turns out that this behavior is shared by a large class of *sub-gaussian random variables*.