Notes for Chapter 2: Concentration of Sums of Independent Random Variables

Gallant Tsao

June 24, 2025

Why Concentration Inequalities?

From previous chapters, the simplest concentration inequality is Chebyshev's Inequality, which is quite general but the bounds can often can be too weak. We can look at the following example:

Example. Toss a fair coin N times. What is the probability that we get at least $\frac{3}{4}$ heads?
Let S_N denote the number of heads, then $S_N \sim \text{Binom}(N, \frac{1}{2})$. We get

$$\mathbb{E}[S_N] = \frac{N}{2}, \text{Var}(S_n) = \frac{N}{4}.$$

Using Chebyshev's Inequality, we get

$$P(S_N \geq \frac{3}{4}N) \leq P(\left|S_N - \frac{N}{2}\right| \geq \frac{N}{4}) \leq \frac{4}{N}.$$

This means probabilistic bound from above converges linearly in N.
However, by using the Central Limit Theorem, we get a very different result: If we let S_N be a sum of independent $Be(\frac{1}{2})$ random variables. Then by the De Moivre-Laplace CLT, the random variable

$$Z_N = \frac{S_N - N/2}{\sqrt{N/4}}$$

converges to the standard normal distribution $N(0,1)$. Then for a large N,

$$P(S_{\geq}\frac{3}{4}N) = P(Z_N \geq \sqrt{N/4}) \approx P(Z \geq \sqrt{N/4})$$

where $Z \sim N(0,1)$. We will use the following proposition:

Proposition (Gaussian tails)**.** *Let $Z \sim N(0,1)$. Then for all $t > 0$,*

$$\frac{t}{t^2+1} \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \leq P(Z \geq t) \leq \frac{1}{t} \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

Proof. The first inequality is proved in exercise 2.2. For the second inequality, by making the change of variables $x = t + y$,

$$\begin{aligned}
P(Z \geq t) &= \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-x^2/2}\ dx \\
&= \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-t^2/2} e^{-ty} e^{-y^2/2}\ dy \\
&\leq \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \int_0^\infty e^{-ty}\ dy \quad (e^{-y^2/2} \leq 1) \\
&= \frac{1}{t} \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.
\end{aligned}$$

□

Therefore the probability of having at least $\frac{3}{4}N$ heads is bounded by

$$\frac{1}{\sqrt{2\pi}}e^{-N/8},$$

which is much better than the linear convergence we had above. However, this reasoning is not rigorous, as the approximation error decays slowly, which can be shown via the CLT below:

Theorem (Berry-Esseen CLT). *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean μ and variance σ^2, and let $S_N = X_1 + \cdots + X_N$, and let*

$$Z_N = \frac{S_N - \mathbb{E}[S_N]}{\sqrt{\mathrm{Var}(S_N)}}.$$

Then for every $N \in \mathbb{N}$ and $t \in \mathbb{R}$ we have

$$|P(Z_N \geq t) - P(Z \geq t)| \leq \frac{\rho}{\sqrt{N}},$$

where $Z \sim N(0,1)$ and $\rho = \mathbb{E}[|X_1 - \mu|^3]/\sigma^3$.