

Lecture 02: Bounding tail distributions of a random variable

*Lecturer: Yuan Zhou**Scribe: Yuan Xie & Yuan Zhou*

Let us consider the unbiased coin flips again. I.e. let the outcome of the i -th coin toss to be a random variables

$$X_i = \begin{cases} +1, & w.p. \frac{1}{2} \\ -1, & w.p. \frac{1}{2} \end{cases}.$$

We assume all coin tosses are independent and let would like to study the sum S of the first n coin tosses,

$$S = \sum_{i=1}^n X_i.$$

In this lecture, we would like the study the probability that S greatly deviates from its mean $\mathbb{E}S = 0$. Specifically, for some parameter t , we would like to estimate the probability $\Pr[S > t]$. Intuitively, we know that such probability should be “small” for large enough t . The goal of this lecture (and part of the next one) is to derive qualitative upper bounds on the tail probability mass parameterized by t .

As we did in the previous lecture, using the Berry-Esseen theorem, we know that

$$|\Pr[S \geq \sqrt{n} \cdot t] - \Pr[G \geq t]| \leq \frac{O(1)}{\sqrt{n}},$$

where $G \sim \mathcal{N}(0, 1)$ is a standard Gaussian. For convenience, we may also use the following informal notation

$$\Pr[S \geq \sqrt{n} \cdot t] = \Pr[G \geq t] \pm \frac{O(1)}{\sqrt{n}}. \quad (1)$$

Using basic calculus, we can estimate that

$$\Pr[G \geq t] = \int_t^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \leq O(1) \cdot e^{-\frac{t^2}{2}}. \quad (2)$$

Now let us fix the parameter $t = 10\sqrt{\ln n}$. Combining (1) and (2), we have

$$\begin{aligned} \Pr[S \geq \sqrt{n} \cdot t] &\leq \Pr[G \geq t] + O\left(\frac{1}{\sqrt{n}}\right) = O\left(\exp\left(-\frac{(10\sqrt{\ln n})^2}{2}\right)\right) + O\left(\frac{1}{\sqrt{n}}\right) \\ &= O\left(\frac{1}{n^{50}}\right) + O\left(\frac{1}{\sqrt{n}}\right) = O\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

We see that the tail mass of the standard Gaussian is only $O\left(\frac{1}{n^{50}}\right)$. However, the error term $O\left(\frac{1}{\sqrt{n}}\right)$ introduced by the Berry-Esseen theorem is much greater. This error term is the main reason that we can't get better results. In the following part of this lecture, we will try various other methods to improve the upper bound.

1 Markov inequality

When we only know the mean of a nonnegative random variable, Markov inequality gives a simple upper bound on the probability that it deviates from its mean.

Theorem 1 (Markov inequality). *Given a random variable X , assume $X \geq 0$. For each parameter $t \geq 1$, we have $\Pr[X \geq t \cdot \mathbb{E}[X]] \leq \frac{1}{t}$.*

Proof. For each $\alpha > 0$, we have

$$\begin{aligned} \mathbb{E}[X] &= \Pr[X \geq \alpha] \cdot \mathbb{E}[X|X \geq \alpha] + \Pr[X < \alpha] \cdot \mathbb{E}[X|X < \alpha] \\ &\geq \Pr[X \geq \alpha] \cdot \alpha + \Pr[X < \alpha] \cdot 0 = \Pr[X \geq \alpha] \cdot \alpha. \end{aligned}$$

Dividing both sides of the inequality by $\alpha > 0$

$$\frac{\mathbb{E}[X]}{\alpha} \leq \Pr[X \geq \alpha]$$

Taking $\alpha = \mathbb{E}[X] \cdot t$ we get the desired bound. \square

Now let us try to apply the Markov inequality to bounding the tail mass of S . Since S is not a nonnegative random variable, we cannot directly apply the inequality. However, note that $S \geq -n$ always holds. We apply the inequality to $T = S + n$ where $\mathbb{E}T = \mathbb{E}S + n = n$. Let $t = 10\sqrt{n \ln n}$. We have

$$\Pr[S \geq t] = \Pr[T \geq t + n] = \Pr\left[T \geq (\mathbb{E}T) \cdot \frac{t + n}{n}\right] \leq \frac{n}{t + n} = \frac{n}{n + 10\sqrt{n \ln n}}.$$

This is a very bad bound – it does not even converge to 0 as n grows!

2 Chebyshev inequality

The Chebyshev inequality not only uses the mean of the random variable, but also needs the variance (or the second moment). Since we have more information about the random variable, we may potentially get better bounds.

Theorem 2 (Chebyshev inequality). Assume that $E[X] = \mu$ and $\text{Var}[X] = \sigma^2 > 0$. For every parameter $t > 0$, we have $\Pr[|X - \mu| \geq t \cdot \sigma] \leq \frac{1}{t^2}$.

Proof. Let $Y = (X - \mu)^2$. We can check that $E[Y] = \sigma^2$ and $Y \geq 0$. Applying Markov inequality, we have

$$\Pr[|X - \mu| \geq t \cdot \sigma] = \Pr[(X - \mu)^2 \geq t^2 \cdot \sigma^2] = \Pr[Y \geq t^2 \cdot E[Y]] \leq \frac{1}{t^2}.$$

□

Now let us go back to the scenario discussed at the beginning of this lecture. We compute that

$$\mu = ES = 0 \quad \text{and} \quad \sigma = \sqrt{\text{Var}[S]} = \sqrt{ES^2} = \sqrt{n}.$$

Therefore

$$\begin{aligned} \Pr[S \geq 10\sqrt{n \ln n}] &\leq \Pr[|S| \geq 10\sqrt{n \ln n}] \\ &= \Pr\left[|S| \geq \sigma \cdot \frac{10\sqrt{n \ln n}}{\sigma}\right] \leq \frac{\sigma^2}{(10\sqrt{n \ln n})^2} = \frac{1}{100 \ln n}. \end{aligned}$$

This bound is still not as good as expected. However, at least it converges to 0 as $n \rightarrow \infty$.

Remark 1. Note that Chebyshev inequality only needs pairwise independence among X_i 's. Specifically, when computing the variance of S , we have

$$\begin{aligned} \text{Var}[S] &= \text{Var}[X_1 + X_2 + \dots + X_n] \\ &= E[(X_1 + X_2 + \dots + X_n)^2] - E[(X_1 + X_2 + \dots + X_n)]^2 = E[(X_1 + X_2 + \dots + X_n)^2] \\ &= \sum_{i=1}^n E[X_i^2] + \sum_{i \neq j} E[X_i X_j] = \sum_{i=1}^n E[X_i^2] = n. \end{aligned}$$

In the penultimate equality, we used the fact that X_i is independent from X_j for $i \neq j$.

3 The fourth moment method

Using the first two moments, we had better bounds than only using the mean of the random variable. Now let us try to extend this method to the fourth moment.

Let us consider $S^4 \geq 0$. By Markov inequality, we have

$$\Pr[S \geq 10\sqrt{n \ln n}] \leq \Pr[S^4 \geq (10\sqrt{n \ln n})^4] \leq \frac{ES^4}{10000n^2 \ln^2 n}. \quad (3)$$

Now let us estimate

$$\begin{aligned} \mathbb{E}[S^4] &= \mathbb{E}\left[\left(\sum_i X_i\right)^4\right] \\ &= \sum_i \mathbb{E} X_i^4 + \sum_i \sum_{j \neq i} \mathbb{E} X_i^2 X_j^2 \cdot \frac{1}{2} \binom{4}{2} + \sum_i \sum_{j \neq i} \mathbb{E} X_i X_j^3 \cdot \binom{4}{1} + \sum_i \sum_{j \neq i} \sum_{k: k \neq i, k \neq j} \mathbb{E} X_i X_j X_k^2 \cdot \binom{4}{2} \\ &\quad + \sum_i \sum_{j \neq i} \sum_{k: k \neq i, k \neq j} \sum_{q: q \neq i, q \neq j, q \neq k} \mathbb{E} X_i X_j X_k X_q. \quad (4) \end{aligned}$$

Fortunately because of independence and $\mathbb{E} X_i = 0$, we have that $\mathbb{E} X_i X_j^3 = \mathbb{E} X_i X_j X_k^2 = \mathbb{E} X_i X_j X_k X_q = 0$. Therefore we can simplify (4) by

$$\mathbb{E}[S^4] = \sum_i \mathbb{E} X_i^4 + \sum_i \sum_{j \neq i} \mathbb{E} X_i^2 X_j^2 \cdot 3 = n + 3n(n-1) \leq 3n^2. \quad (5)$$

Combining (3) and (5), we get

$$\Pr[S \geq 10\sqrt{n \ln n}] \leq \frac{3n^2}{10000n^2 \ln^2 n} = \frac{3}{10000 \ln^2 n}.$$

This is a better bound than what we get from Chebyshev.

Remark 2. *When using the fourth moment method, we only used the independence among every quadruple of random variables. Therefore the bound works for 4-wise independent random variables too.*

Remark 3. *We can extend this method to considering S^{2k} for positive integer k 's, and picking the k that optimizes the upper bound. However, this plan would lead to the painful estimation of $\mathbb{E} S^{2k}$. We will use a slightly different method to get better upper bounds.*

4 The “Chernoff method”

Instead of S^{2k} , let us consider the function $e^{\lambda S}$ for some positive parameter λ .

Since e^x is a monotonically increasing function, we have

$$\Pr[S \geq 10\sqrt{n \ln n}] = \Pr[\lambda S \geq 10\lambda\sqrt{n \ln n}] \leq \Pr[e^{\lambda S} \geq e^{10\lambda\sqrt{n \ln n}}].$$

By Markov inequality (also checking that $e^{\lambda S} > 0$, we have

$$\Pr[e^{\lambda S} \geq e^{10\lambda\sqrt{n \ln n}}] \leq \frac{\mathbb{E} e^{\lambda S}}{e^{10\lambda\sqrt{n \ln n}}}. \quad (6)$$

Now it remains to upper bound $\mathbb{E} e^{\lambda S}$. We have

$$\mathbb{E} e^{\lambda S} = \mathbb{E} e^{\lambda \sum_i X_i} = \mathbb{E} \prod_i e^{\lambda X_i} = \prod_i \mathbb{E} e^{\lambda X_i}. \quad (7)$$

Note that in the last equality, we used the full independence among all X_i 's.

On the other hand, by the distribution of X_i , we have

$$\begin{aligned} \mathbb{E} e^{\lambda X_i} &= \frac{1}{2} e^\lambda + \frac{1}{2} e^{-\lambda} \\ &= \frac{1}{2} \left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \frac{\lambda^4}{4!} + \dots \right) + \frac{1}{2} \left(1 - \lambda + \frac{\lambda^2}{2!} - \frac{\lambda^3}{3!} + \frac{\lambda^4}{4!} - \dots \right) \\ &\hspace{15em} \text{(Taylor expansion)} \\ &= 1 + \frac{\lambda^2}{2!} + \frac{\lambda^4}{4!} + \frac{\lambda^6}{6!} + \dots \leq e^{\lambda^2/2}. \end{aligned}$$

Getting back to (7), we have

$$\mathbb{E} e^{\lambda S} \leq \prod_i e^{\lambda^2/2} = e^{n\lambda^2/2}.$$

Combining this with (6), we have

$$\Pr[e^{\lambda S} \geq e^{10\lambda\sqrt{n\ln n}}] \leq e^{n\lambda^2/2 - 10\lambda\sqrt{n\ln n}}. \quad (8)$$

Picking $\lambda = 10\sqrt{\frac{\ln n}{n}}$, we minimize the right-hand side of (8) and get our desired upper bound

$$\Pr[e^{\lambda S} \geq e^{10\lambda\sqrt{n\ln n}}] \leq e^{50\ln n - 100\ln n} = \frac{1}{n^{50}}.$$

In the beginning of the next lecture, we are going to extend this method to more general random variables and general thresholds, and go through the proof the famous Chernoff bound.