Contents

1 Introduction

1.1 Basic Inequalities

Theorem 1.1.1 (Markov's inequality)**.** *For a random variable X with $\mathbf{P}\{X < 0\} = 0$ and $t > 0$, we have*

$$\mathbf{P}\{X \geq t\} \leq \frac{\mathbf{E}\, X}{t}.$$

It follows that for a non-decreasing function φ which only takes non-negative values,

$$\mathbf{P}\{X \geq t\} = \mathbf{P}\{\varphi(X) \geq \varphi(t)\} \leq \frac{\varphi(X)}{\varphi(t)}.$$

Proof. In the first place, note that

$$\begin{aligned} X &= X \cdot \mathbb{1}_{X \geq t} + X \cdot \mathbb{1}_{X < t} \\ &\geq t \cdot \mathbb{1}_{X \geq t} + 0, \end{aligned}$$

and thus,

$$\mathbf{E}\, X \geq t \cdot \mathbf{E}\, \mathbb{1}_{X \geq t} = t \cdot \mathbf{P}\{X \geq t\}.$$

For the second statement, apply the same argument on the random variable $Y := \varphi(X)$ and the constant $s := \varphi(t)$. □

Theorem 1.1.2 (Chebyshev's inequality)**.** *For $t > 0$ and a random variable X with mean $\mu = \mathbf{E}\, X$ and variance $\sigma^2 = \mathbf{Var}\, X$, then*

$$\mathbf{P}\{|X - \mu| \geq t\} \leq \sigma^2 t^{-2}.$$

Proof. Applying Markov's inequality with $\varphi : x \mapsto x^2$ we obtain,

$$\mathbf{P}\{|X - \mu| \geq t\} = \mathbf{P}\{|X - \mu|^2 \geq t^2\} \leq \frac{\mathbf{E}\left[(X - \mu)^2\right]}{t^2} = \sigma^2 t^{-2}.$$

□

Theorem 1.1.3 (Jensen's inequality)**.** *For any real valued random variable X and convex function φ*

$$\varphi(\mathbf{E}\, X) \leq \mathbf{E}\, \varphi(X)$$

1.2 Why bother?

The concentration inequalities are used to obtain information on how a random variable is distributed at some specific places of its domain. In the most common scenarios, these inequalities will be used to quantify how concentrated a random variable at its tails, for example,

$$\mathbf{P}\{|X-\mu| \geq t\} < f(t) << 1.$$

A concentration inequality is specially useful when this probability cannot be calculated at a low computational cost or estimated with high precision. The following will illustrate a case where using concentration inequalities achieves the best results.

1.2.1 Coin Tossing

A coin tossing game is fair if the chances of winning are equal to the chances of losing. We can verify from a sample of N games that the game is not rigged if the number of heads in the sample is not very distant from the average $N/2$. However, there's a chance that one may classify the coin as rigged, even when the coin is fair. By the *Law of Large Numbers*, we know that the larger the sample, the less likely it is to obtain a false positive. But let's ask ourselves how fast this probability converges to 0.

Let $S_N \sim \mathrm{Bi}(N, 1/2)$ denote the number of heads in a fair coin tossing game. Then,

$$\mu = \mathbf{E}\, S_N = \frac{N}{2}, \qquad \sigma^2 = \mathbf{Var}\, S_N = \frac{N}{4}.$$

For a fixed $\varepsilon > 0$, we may classify a coin tossing game as rigged if, after N trials, the ratio of heads vs tails in the sample is greater than $[1+\varepsilon : 1-\varepsilon]$, or similarly,

$$S_N \geq \mu + \frac{\varepsilon}{2}N = \frac{1+\varepsilon}{2}N.$$

Using the Chebyshev inequality 1.1.2, we assert that

$$\mathbf{P}\left\{S_N \geq \mu + \frac{\varepsilon}{2}N\right\} \leq \mathbf{P}\left\{|S_N - \mu| \geq \frac{\varepsilon}{2}N\right\} \leq \sigma^2 \frac{4}{\varepsilon^2 N^2} = \frac{1}{\varepsilon^2 N}.$$

Therefore, the probability of bad events tends to 0 at least linearly with the number of games.

1.2.2 Central Limit Theorem

The proof of the following theorems can be found in (ref)

Theorem 1.2.1. *Let X_i be a i.i.d. sample. Let $S_N = \sum_{i=1}^N X_i$, with mean $\mu = \mathbf{E}\, S_N$ and variance $\sigma^2 = \mathbf{Var}\, S_N$. If*

$$Z_N = \frac{S_N - N \cdot \mathbf{E}\, X_i}{\sqrt{N \cdot \mathbf{Var}\, X_i}} = \frac{S_N - \mu}{\sqrt{N}\sigma},$$

then,

$$Z_N \to Z \sim \mathcal{N}(0,1), \text{ in distribution.}$$

$\square$

Theorem 1.2.2 (Tails of the Normal Distribution). *Let $Z \sim \mathcal{N}(0,1)$, for $t > 0$ we have*

$$\left(\frac{1}{t} - \frac{1}{t^3}\right) \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) \leq \mathbf{P}\{Z \geq t\} \leq \frac{1}{t} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right).$$

$\square$

With that in mind, we might naively assume that better bounds can be obtained by using the previous theorem. For a large enough N we can say that for the coin tossing,

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

$$\implies \mathbf{P}\left\{S_N \geq \frac{1+\varepsilon}{2} N\right\} = \mathbf{P}\left\{Z_N \geq \varepsilon\sqrt{N}\right\} \sim \mathbf{P}\left\{Z \geq \varepsilon\sqrt{N}\right\}.$$

However, this raises the question of whether we can draw the following conclusion from Theorem 1.2.2:

$$\mathbf{P}\left\{S_N \geq \frac{1+\varepsilon}{2} N\right\} \leq \frac{1}{\varepsilon\sqrt{N}} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-\varepsilon^2 \cdot N}{2}\right).$$

Unfortunately, the answer is no. The following theorem will show why.

Theorem 1.2.3 (Convergence Rate for Central Limit Theorem). *For Z_N, Z in Theorem 1.2.1, we have:*

$$|\mathbf{P}\{Z_N \geq t\} - \mathbf{P}\{Z \geq t\}| \in O(\tfrac{1}{\sqrt{N}}).$$

$\square$

Since the approximation error is greater than the bound, the previous results cannot be taken into account.

In the context of coin tossing, this may not matter at all because the linear bound obtained using Chebyshev's inequality indicates that the probability of wrongly classifying a fair coin as a rigged coin converges at least linearly to zero. Even the Central Limit Theorem shows in a less precise way this convergence. However, for some specific problems in statistics, these basic tools are not precise enough to solve them. In the following chapters, we will show some examples were better crafted strategies are needed in order to get bounds to the tails of the random variables.

2 Exponential Inequalities

2.1 Chernoff-Okamoto Inequalities

Applying Markov's Inequality to $Y = e^{uX}$, we can assert that

$$\mathbf{P}\{X \geq \lambda + t\} \leq e^{-u(\lambda+t)} \mathbf{E}\, e^{uX} = e^{-u(\lambda+t)}(1 - p + pe^u)^n.$$

The right hand equation is minimized when,

$$e^u = \frac{\lambda + t}{(n - \lambda - t)} \cdot \frac{1-p}{p}.$$

Therefore, for $0 \leq t \leq n - \lambda$,

$$\mathbf{P}\{X \geq \lambda + t\} \leq \left(\frac{\lambda}{\lambda + t}\right)^{\lambda+t} \left(\frac{n-\lambda}{n-\lambda-t}\right)^{n-\lambda-t} \tag{2.1}$$

Theorem 2.1.1. *Let X be random variable with the binomial distribution* $\mathrm{Bi}(n,p)$ *with $\lambda := np = \mathbf{E}\, X$, then for $t \geq 0$,*

$$\mathbf{P}\{X \geq \lambda + t\} \leq \exp\left(-\frac{t^2}{2(\lambda + t/3)}\right) \tag{2.2}$$

$$\mathbf{P}\{X \leq \lambda - t\} \leq \exp\left(-\frac{t^2}{2\lambda}\right) \tag{2.3}$$

Used in: Theorem **??**

Proof. (**TODO** I've already written the proof on paper) □

2.2 Hoeffding-Bernstein inequalities

Theorem 2.2.1. *Let $\|f\|_\infty < c$, $\mathbf{E}\, f(X_1, \ldots, X_m) = 0$ and $\sigma^2 = \mathbf{E}\, f^2(X_1, \ldots, X_m)$. Then for any $t > 0$,*

$$\mathbf{P}\{U_m^n(f, P) > t\} \leq \exp\left(\frac{\frac{n}{m}t^2}{2\sigma^2 + \frac{2}{3}ct}\right) \tag{2.1}$$

Used in: Theorem **??**

Proof. Proposition 2.3(a) M.A. Arcones, E. Gine, Limit theorems for U-processes, Ann. Probab. 21 (1993) 14941542

https://sci-hub.se/https://www.jstor.org/stable/2244585 □