# Online Learning and Decision Making

Tianjiao Nie

September 16, 2025

## Contents

## 1 Introduction

Consider a dynamic pricing process with two prices $\{\alpha, \beta\}$. The model is $d = D(p) + \epsilon$ where $\mathbb{E}[\epsilon] = 0$. Consider the following simple strategy (called *exploration-then-commit*):

1. offer $p = \alpha$ for $m$ times, and observe $d_1, \ldots, d_m$

2. offer $p = \beta$ for $m$ times, and observe $d_{m+1}, \ldots, d_{2m}$

3. estimate $\widehat{D}(\alpha) = (d_1 + \cdots + d_m)/m$ nd $\widehat{D}(\beta) = (d_{m+1} + \cdots + d_{2m})/m$

4. commit the price $\arg\max_{p \in \{\alpha, \beta\}} p \cdot \widehat{D}(p)$ for the rest $T - 2m$ times.

An immediate question is how to choose a suitable $m$.
We would like an estimate of the form

$$\Pr\left[\left|D(\alpha) - \widehat{D}(\alpha)\right| \leq \tau\right] \geq 1 - \delta.$$

Recall that $\tau$ is called the *confidence radius* and $1 - \delta$ is called the *confidence level*.

## A Concentration Inequalitites

**Lemma A.1** (Berry–Esseen). Let $X_1, \ldots, X_m$ be i.i.d. random variables such that $\mathbb{E}[X_i] = 0$ and $\sum_{i=1}^{m} \mathrm{Var}[X_i] = 1$. Let $S = X_1 + \cdots + X_m$. We have $\mathbb{E}[S] = 0$ and $\mathrm{Var}[S] = 1$. Then for every $u \in \mathbb{R}$,

$$\left|\Pr[S \leq u] - \Pr_{Z \sim \mathcal{N}(0,1)}[Z \leq u]\right| \leq C\beta$$

where $C > 0$ is a universal constant and $\beta = \sum_{i=1}^{m} \mathbb{E}\left[|X_i|^3\right]$ is the third momentum.

**Example A.2.** Consider $2m$ fair coins. Then

$$\Pr[\#H = \#T] = \frac{(2m)!}{m!m!2^{2m}} \sim \frac{\sqrt{2\pi \cdot 2m}(2m/e)^{2m}}{\left[\sqrt{2\pi m}(m/e)^m\right]^2 2^{2m}} = \frac{1}{\sqrt{\pi m}}.$$

**Lemma A.3** (Markov). Let $X \geq 0$ be a random variable with $\mathbb{E}[X] = \mu$. Then for $t > 0$,

$$\Pr[X \geq t\mu] \leq 1/t.$$

*Proof.* This follows from $\mathbb{E}[X] \geq \Pr[X \geq t] \cdot t + \Pr[X < t] \cdot 0 = t\Pr[X \geq t]$. $\square$

**Lemma A.4** (Chebyshev). Let $X$ be a random variable with $\mathbb{E}[X] = \mu$ and $\mathrm{Var}[X] = \sigma^2$. Then for $t > 0$,

$$\Pr\left[|X - \mu| \geq t\sigma\right] \leq 1/t^2.$$