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 $d_1, \dots, d_m$
2. offer $p = \beta$ for m times, and observe $d_{m+1}, \dots, d_{2m}$
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

$$\Pr\Big[\big|D(\alpha) - \widehat{D}(\alpha)\big| \le \tau\Big] \ge 1 - \delta.$$

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

A Concentration Inequalitites

Lemma A.1 (Berry–Esseen)**.** Let $X_1, \dots, X_m$ 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 $u \in \mathbb{R}$,

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

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

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

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

where we used Stirling's approximation to estimate factorials.

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

$$\Pr[X \ge t\mu] \le 1/t.$$

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

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

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

Proof. Note that $\Pr\big[|X - \mu| \geq t\sigma\big] = \Pr\big[|X - \mu|^2 \geq t^2\sigma^2\big]$ and then we apply Markov's inequality. □