Minimax Lower Bound for Gaussian Bandits

1 Gaussian Tail Bound and Hypothesis Testing

Let $X_1, \ldots, X_n$ be i.i.d. Gaussian random variables with variance 1 and mean $\mu \in \{0, \Delta\}$. Define the sample mean

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^{n} X_t.$$

Then

$$\hat{\mu} \sim \mathcal{N}\left(\mu, \frac{1}{n}\right).$$

In particular, under $\mu = 0$,

$$\mathbb{P}\left(\hat{\mu} \geq \frac{\Delta}{2}\right) = \mathbb{P}\left(Z \geq \frac{\sqrt{n}\,\Delta}{2}\right), \quad Z \sim \mathcal{N}(0,1).$$

A sharp Gaussian tail bound (Mills-type inequality) yields that for all $x \geq 0$,

$$\frac{e^{-x^2}}{x + \sqrt{x^2+2}} \;\leq\; \int_x^\infty e^{-t^2}\,dt \;\leq\; \frac{e^{-x^2}}{x + \sqrt{x^2 + 4/\pi}}.$$

Applying this bound with $x = \frac{\sqrt{n}\,\Delta}{2\sqrt{2}}$ gives

$$\begin{aligned} \frac{1}{\sqrt{n\Delta^2} + \sqrt{n\Delta^2 + 16}} \sqrt{\frac{8}{\pi}}\, e^{-n\Delta^2/8} \;&\leq\; \mathbb{P}\left(\hat{\mu} \geq \frac{\Delta}{2}\right) \\ &\leq\; \frac{1}{\sqrt{n\Delta^2} + \sqrt{n\Delta^2 + 32/\pi}} \sqrt{\frac{8}{\pi}}\, e^{-n\Delta^2/8}. \end{aligned} \tag{1}$$

Key message. The error probability scales as

$$\mathbb{P}(\text{error}) = \exp\bigl(-\Theta(n\Delta^2)\bigr),$$

so the distinguishability threshold is $n\Delta^2 \approx 1$.

2 Bandit Setup

Consider a k-armed Gaussian bandit with unit variance. Let π be an arbitrary policy and $T_j(n)$ the number of pulls of arm j up to time n.

Define two environments:

- Environment ν with mean vector

$$\mu = (\Delta, 0, 0, \dots, 0),$$

so arm 1 is optimal.

- Environment ν' obtained by modifying a single arm:

$$\mu'_j = \begin{cases} \mu_j, & j \neq i, \\ 2\Delta, & j = i, \end{cases}$$

where

$$i := \arg\min_{j>1} \mathbb{E}[T_j(n)].$$

In ν', arm i is optimal.

3 Expected Regret Decomposition

In environment ν,

$$R_n(\pi, \nu) = (n - \mathbb{E}[T_1(n)])\Delta.$$

In environment ν', the optimal mean is $\mu^\star = 2\Delta$, hence

$$\begin{aligned} R_n(\pi, \nu') &= \Delta\, \mathbb{E}_{\nu'}[T_1(n)] + \sum_{j \notin \{1,i\}} 2\Delta\, \mathbb{E}_{\nu'}[T_j(n)] \\ &\geq \Delta\, \mathbb{E}_{\nu'}[T_1(n)]. \end{aligned} \tag{2}$$

4 Choice of Δ

Since there are $k-1$ suboptimal arms in ν,

$$\sum_{j>1} \mathbb{E}[T_j(n)] \leq n \quad \Rightarrow \quad \mathbb{E}[T_i(n)] \leq \frac{n}{k-1}.$$

Motivated by the Gaussian testing threshold (1), we choose

$$\Delta = \sqrt{\frac{1}{\mathbb{E}[T_i(n)]}} \geq \sqrt{\frac{k-1}{n}}.$$

With this choice, the policy cannot reliably distinguish ν from ν', and heuristically

$$\mathbb{E}[T_1(n)] \approx \mathbb{E}_{\nu'}[T_1(n)].$$

5 Case Analysis

Case 1: $\mathbb{E}[T_1(n)] < \frac{n}{2}$. Using the regret in ν,

$$R_n(\pi, \nu) \geq \frac{n}{2}\Delta \geq \frac{1}{2}\sqrt{n(k-1)}.$$

Case 2: $\mathbb{E}[T_1(n)] \geq \frac{n}{2}$. Using (2),

$$R_n(\pi, \nu') \geq \Delta\, \mathbb{E}_{\nu'}[T_1(n)] \approx \Delta\, \mathbb{E}[T_1(n)] \geq \frac{1}{2}\sqrt{n(k-1)}.$$

6 Minimax Lower Bound

In both cases,

$$\max\{R_n(\pi, \nu), R_n(\pi, \nu')\} \;\geq\; c\sqrt{n(k-1)}$$

for a universal constant $c > 0$. Therefore,

$$R_n^*(\mathcal{E}^k) \;\geq\; c\sqrt{nk},$$

which establishes the minimax lower bound (up to constants).