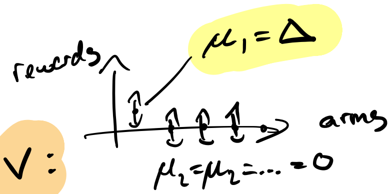
7.2.2. Constructing worst-case environments

Fix π

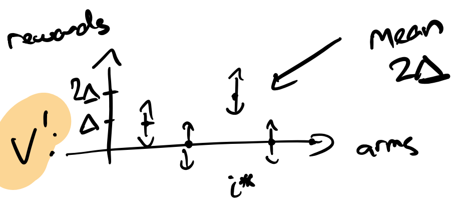
ν :



reward for arm i has Gaussian distribution with mean μ_i and variance 1.

• Define $i^* := \underset{i \in [k] \setminus \{1\}}{\operatorname{argmin}} \mathbb{E}[T_n(i)]$ ← the arm that is played the least

ν' :



Regret: $R_n(\pi, \nu) = \Delta \sum_{i=2}^{k} \mathbb{E}[T_i(n)]$

$$= \Delta(n - \mathbb{E}[T_1(n)]) \qquad ①$$

$$R_n(\pi, \nu') = \Delta \mathbb{E}'[T_1(n)] + 2\Delta \sum_{i \in [k] \setminus \{1, i^*\}} \mathbb{E}'[T_i(n)] \qquad ②$$

$$\geq \Delta \mathbb{E}'[T_1(n)]$$

$$\mathbb{E}[T_1(n)] \leq \frac{n}{2} P[T_1(n) < \frac{n}{2}] + n P[T_1(n) > \frac{n}{2}]$$

$$= \frac{n}{2}(1 + P[T_1(n) \geq \frac{n}{2}])$$

$$\mathbb{E}'[T_1(n)] \geq \frac{n}{2} P'\left[T_1(n) \geq \frac{n}{2}\right]$$

Plugging this into ① and ②:

$$R_n(\pi, \nu) + R_n(\pi, \nu') \geq \frac{n\Delta}{2}\left(P\left[T_1(n) < \frac{n}{2}\right] + P'\left[T_1(n) \geq \frac{n}{2}\right]\right)$$

7.2.3 Lower-bounding the regret

Bretagnolle-Huber inequality:

Lemma: Let p and q be two pdfs for X taking values in $\mathcal{X}$. For any $A \subset \mathcal{X}$, we have

$$p(A) + q(A^c) \geq \frac{1}{2} e^{-D(p\|q)}.$$

Proof:

$$p(A) + q(A^c) = \int_A p(x)\,dx + \int_{A^c} q(x)\,dx$$

$$\geq \int_{\mathcal{X}} \min\{p(x), q(x)\}\,dx \qquad (*)$$

Using Cauchy-Schwartz:

$$\left(\int_{\mathcal{X}} \sqrt{p(x)q(x)}\right)^2 = \left(\int_{\mathcal{X}} \sqrt{\min\{p(x), q(x)\} \max\{p(x), q(x)\}}\right)^2$$

$$\leq \int_{\mathcal{X}} \min\{p(x), q(x)\} \cdot \int_{\mathcal{X}} \max\{p(x), q(x)\}$$

$$\leq 2 \int \min\{p(x), q(x)\}$$

Combining with (*):

$$p(A) + q(A^c) \geq \frac{1}{2}\left(\int_x \sqrt{p(x)q(x)}\right)^2$$

$$= \frac{1}{2}\exp\left(2\log\int_x \sqrt{p(x)q(x)}\,dx\right)$$

$$= \frac{1}{2}\exp\left(2\log\int_x p(x)\sqrt{\frac{q(x)}{p(x)}}\,dx\right)$$

Jensen's inequality

$$\geq \frac{1}{2}\exp\left(2\int_x p(x)\log\sqrt{\frac{q(x)}{p(x)}}\,dx\right)$$

$$= \frac{1}{2}\exp\left(-\int_x p(x)\log\frac{p(x)}{q(x)}\,dx\right)$$

$$= \frac{1}{2}\exp\left(-D(p\|q)\right) \qquad \square$$

$$\Rightarrow \quad \underline{R_n(\pi, \nu) + R_n(\pi, \nu') \geq \frac{n\Delta}{4} e^{-D(p\|p')}}$$

Lemma: $D(p\|p') = \sum_{i=1}^{k} \mathbb{E}[T_i(n)]\, D(p_i\|p_i')$

relative entropy

Proof: ommitted

In our case: $D(p\|p') = \mathbb{E}[T_{i^*}(n)] \cdot \underbrace{\frac{(2\Delta)^2}{2}}_{\text{relative entropy between } \mathcal{N}(\cdot,0,1) \text{ and } \mathcal{N}(\cdot,2\Delta,1)}$

$$< \frac{n}{\ } \cdot 2\Delta^2$$

$- \quad k-1 \quad -$

$$\Rightarrow \quad R_n(\pi, \nu) + R_n(\pi, \nu') \geq \frac{n\Delta}{4} e^{-\frac{n}{k-1} 2\Delta^2}$$

$$= \sqrt{n(k-1)} \cdot \underbrace{\frac{1}{8} e^{-\frac{1}{2}}}_{\geq \frac{2}{27}}$$

Choose Δ!

$$\Delta = \sqrt{\frac{k-1}{4n}}$$

$$\geq \frac{2}{27} \sqrt{n(k-1)}$$

$$\Rightarrow \quad \max_{\nu} R_n(\pi, \nu) \geq \frac{1}{27} \sqrt{n(k-1)}$$

for all policies π!

This shows the theorem!