# Stat 260/CS 294-102. Learning in Sequential Decision Problems.

**Peter Bartlett**

1. Minimax regret bounds
   - Upper bounds: worst case over $\Delta_j$.
   - Lower bounds.

## Pseudo-regret

Recall

$$\overline{R}_n(P_\theta) = \max_{j^*=1,\ldots,k} \mathbb{E}\left[\sum_{t=1}^{n} X_{j^*,t} - \sum_{t=1}^{n} X_{I_t,t}\right] = n\mu^* - \mathbb{E}\sum_{t=1}^{n} X_{I_t,t}.$$

We have seen pseudo-regret bounds for a particular strategy, of the form: For all reward distributions on $[0,1]$ and all $n$,

$$\overline{R}_n(P) = \sum_{j=1}^{k} \Delta_j \mathbb{E} T_j(n) \le c_1 \sum_{j=1}^{k} \frac{\Delta_j \log n}{d(\mu_j, \mu^*)} + c_2(\mu_1, \ldots, \mu_k),$$

where $d$ is the KL-divergence between Bernoulli distributions with the given expectations. These are obtained by showing that

$$\Delta_j \mathbb{E} T_j(n) \le \frac{c_1 \Delta_j \log n}{d(\mu_j, \mu^*)} + c_2(\mu_1, \ldots, \mu_k).$$

# Minimax upper bounds

These bounds get worse as the $\Delta_j$ get smaller. We can also obtain regret bounds that are independent of the $\Delta_j$, but the rate is worse.

**Theorem:** If a particular strategy satisfies:
for all reward distributions on $[0, 1]$ and all $n$,

$$\mathbb{E}T_j(n) \leq c_1 \frac{\log n}{d(\mu_j, \mu^*)} + \frac{c_2}{\Delta_j^2}.$$

then for all $n$,

$$\sup_P \overline{R}_n(P) \leq \sqrt{kn\left(\frac{c_1}{2}\log n + c_2\right)}.$$

## Minimax upper bounds

We know that, for a fixed distribution, we can achieve a much better regret rate (logarithmic in $n$), but the constant in that rate depends on the distribution. This bound holds uniformly across all distributions. It's a *minimax* bound:

$$\min_S \max_P \overline{R}_n(P) \leq \sqrt{kn\left(\frac{c_1}{2}\log n + c_2\right)},$$

where the min is over strategies.

These are also called *distribution-free* regret bounds.

(Note: the $c_2/\Delta_j^2$ term was unimportant when we were treating the $\Delta_j$ as constants. But its dependence on $\Delta_j$ is important here.)

# Minimax upper bounds: proof

Pinsker's inequality ($d(\mu_j, \mu^*) \geq 2\Delta_j^2$) implies

$$\Delta_j \mathbb{E} T_j(n) \leq \frac{1}{\Delta_j} \left( \frac{c_1}{2} \log n + c_2 \right).$$

If we define $p_j = \mathbb{E} T_j(n)/n$, so that $\sum_j p_j = 1$, then

$$\begin{aligned} \overline{R}_n(P) &= \sum_{j=1}^{k} \Delta_j \mathbb{E} T_j(n) \\ &\leq \sum_{j=1}^{k} \min \left\{ \frac{1}{\Delta_j} \left( \frac{c_1}{2} \log n + c_2 \right), p_j n \Delta_j \right\}. \end{aligned}$$

## Minimax upper bounds: proof

The minimum is maximized for

$$\frac{1}{\Delta_j}\left(\frac{c_1}{2}\log n + c_2\right) = p_j n \Delta_j,$$

and solving for $\Delta_j$ gives

$$np_j\Delta_j = \sqrt{np_j\left(\frac{c_1}{2}\log n + c_2\right)}.$$

(And the two terms in the minimum are monotonically increasing and decreasing in $\Delta_j$, so if this choice of $\Delta_j$ is impossible—e.g., $\Delta_j > 1$—then the minimum is only smaller.)

# Minimax upper bounds: proof

Thus,

$$\begin{aligned}
\overline{R}_n(P) &\leq \sqrt{n\left(\frac{c_1}{2}\log n + c_2\right)} \sum_{j=1}^{k} \sqrt{p_j} \\
&\leq \sqrt{n\left(\frac{c_1}{2}\log n + c_2\right)} \left(\sum_{j=1}^{k} p_j\right)^{1/2} \left(\sum_{j=1}^{k} 1\right)^{1/2} \\
&= \sqrt{kn\left(\frac{c_1}{2}\log n + c_2\right)},
\end{aligned}$$

by Cauchy-Schwarz.

## Minimax lower bound

**Theorem:** Let $\mathcal{P}$ be the set of all Bernoulli reward distributions. Then for all $n$,

$$\inf_{\text{strategies}} \sup_{P \in \mathcal{P}^k} \overline{R}_n(P) \geq \frac{1}{18} \min\{\sqrt{nk}, n\}.$$

Note the order of quantifiers: fix any strategy, then for all $n$, there is a reward distribution for which the regret is $\Omega(\sqrt{nk})$. On the other hand, we know that there are strategies so that for any reward distributions, the regret grows like $O(\log n)$. The lower bound is saying that the envelope of all of these regret curves must be $\Omega(\sqrt{nk})$.

# Minimax lower bound: intuition

After $n$ rounds, some arm has not been pulled more than $n/k$ times. For that arm, the deviations in the sample averages are of the order of $\sqrt{k/n}$, so we cannot hope to identify the best arm on a finer scale than this. So choosing a distribution so that the best arm is only $\sqrt{k/n}$ better than the others, the regret should be roughly $n\sqrt{k/n} = \sqrt{kn}$.

# Minimax lower bound: proof

We'll use the probabilistic method: randomly choose the reward distributions and show that, for any strategy, under that random choice,

$$\mathbb{E}\overline{R}_n(P) \geq \frac{1}{18}\min\{\sqrt{nk}, n\}.$$

This implies that, for that strategy, there must be a reward distribution that incurs at least that regret.

# Minimax lower bound: proof

Rewards:

$$\mu^* = \frac{1}{2} + \epsilon, \qquad \mu_j = \frac{1}{2} \qquad \text{for } j \neq j^*.$$

(We'll choose $\epsilon$ later.)

Choose index $j^*$ uniformly at random.

Fix a strategy. Let $\mathbb{P}_{j^*}$ denote the distribution of the sequence of rewards $Y_t = X_{I_t,t}$ (and the expectation under that distribution) with the fixed strategy and the choice of index $j^*$.

# Minimax lower bound: proof

$$\frac{1}{k}\sum_{j^*=1}^{k}\overline{R}_n(\mathbb{P}_{j^*}) = \frac{1}{k}\sum_{j^*=1}^{k}\mathbb{P}_{j^*}\sum_{j=1}^{k}\Delta_j T_j(n)$$

$$= \frac{\epsilon}{k}\sum_{j^*=1}^{k}\mathbb{P}_{j^*}\sum_{j\neq j^*}T_j(n)$$

$$= \epsilon\left(n - \frac{1}{k}\sum_{j^*=1}^{k}\mathbb{P}_{j^*}T_{j^*}(n)\right).$$

## Minimax lower bound: proof

Let $\mathbb{P}$ denote the distribution of the sequence of rewards $Y_t = X_{I_t,t}$ (and the expectation under that distribution) with the fixed strategy, when the rewards $X_{j,t}$ have $\mu_j = 1/2$ for *all* $j$. Then

$$
\begin{aligned}
\mathbb{P}_{j^*} T_{j^*}(n) &\le \mathbb{P} T_{j^*}(n) + n D_{TV}\left(\mathbb{P}, \mathbb{P}_{j^*}\right) && \text{(see (1) below)} \\
&\le \mathbb{P} T_{j^*}(n) + n \sqrt{\frac{1}{2} D_{KL}\left(\mathbb{P}, \mathbb{P}_{j^*}\right)} && \text{(Pinsker (2))} \\
&= \mathbb{P} T_{j^*}(n) + \frac{n}{2} \sqrt{\log \frac{1}{1-4\epsilon^2} \mathbb{P} T_{j^*}(n)} && \text{(chain rule (3)).}
\end{aligned}
$$

## Minimax lower bound: proof

Notice that $\sum_{j^*=1}^{k} \mathbb{P}T_{j^*}(n) = n$:

$$
\begin{aligned}
&\frac{1}{k}\sum_{j^*=1}^{k} \mathbb{P}_{j^*}T_{j^*}(n) \\
&\leq \frac{1}{k}\sum_{j^*=1}^{k} \mathbb{P}T_{j^*}(n) + \frac{n}{2k}\sum_{j^*=1}^{k}\sqrt{\log\frac{1}{1-4\epsilon^2}\mathbb{P}T_{j^*}(n)} \\
&\leq \frac{n}{k} + \frac{n}{2}\sqrt{\log\frac{1}{1-4\epsilon^2}\frac{1}{k}\sum_{j^*=1}^{k}\mathbb{P}T_{j^*}(n)} \qquad \text{(Jensen)} \\
&= \frac{n}{k} + \frac{n}{2}\sqrt{\frac{n}{k}\log\frac{1}{1-4\epsilon^2}}.
\end{aligned}
$$

# Minimax lower bound: proof

Combining,

$$\frac{1}{k}\sum_{j^*=1}^{k}\overline{R}_n(\mathbb{P}_{j^*}) \geq \epsilon n\left(1-\frac{1}{k}-\frac{1}{2}\sqrt{\frac{n}{k}\log\frac{1}{1-4\epsilon^2}}\right).$$

Since $\log(1-x)$ is concave, the line between two points on its graph lies below the graph: $\log(1-x) \geq -\log(1-c)x/c$ for $0 \leq x \leq c$. So $\log(1/(1-4\epsilon^2)) \geq c\epsilon^2$. Picking $\epsilon = \min(\sqrt{k/n}, 1)/4$ gives

$$\frac{1}{k}\sum_{j^*=1}^{k}\overline{R}_n(\mathbb{P}_{j^*}) \geq \frac{1}{18}\min\left\{\sqrt{kn}, n\right\}.$$

## (1) $D_{TV}$ and changes of expectation

**Lemma:** If $\sup_x f(x) - \inf_x f(x) = 1$,

$$|Pf - Qf| \leq D_{TV}(P, Q).$$

$$\begin{aligned} Pf - Qf &= \int f\, dP - \int f\, dQ \\ &\leq \int 1\left[\frac{dP}{d(P+Q)} > \frac{dQ}{d(P+Q)}\right] d(P-Q) \\ &= D_{TV}(P, Q). \end{aligned}$$

# (2) Pinsker's inequality

**Lemma:**

$$D_{KL}(P, Q) \geq 2D_{TV}(P, Q)^2.$$

## (3) Chain rule for KL-divergence

**Lemma:** For $\mathbb{P}$ and $\mathbb{Q}$ distributions of sequences $Y_1, \ldots, Y_n$,

$$D_{KL}(\mathbb{P}, \mathbb{Q}) = \mathbb{P} \sum_{t=1}^{n} D_{KL}\left(\mathbb{P}(Y_t | Y^{t-1}), \mathbb{Q}(Y_t | Y^{t-1})\right).$$

# (3) Chain rule for KL-divergence

*Proof.*

$$
\begin{aligned}
D_{KL}(\mathbb{P}, \mathbb{Q}) &= \int \log \frac{d\mathbb{P}}{d\mathbb{Q}}(y^n)\, d\mathbb{P}(y^n) \\
&= \int \int \log \left( \frac{d\mathbb{P}(x_n | y^{n-1})}{d\mathbb{Q}(x_n | y^{n-1})}(y_n) \right. \\
&\qquad\qquad \left. \times \frac{d\mathbb{P}(x^{n-1})}{d\mathbb{Q}(x^{n-1})}(y^{n-1}) \right) \, d\mathbb{P}(y_n | y^{n-1})\, d\mathbb{P}(y^{n-1}) \\
&= \int D_{KL}\left(\mathbb{P}(y_n | y^{n-1}), \mathbb{Q}(y_n | y^{n-1})\right) \, d\mathbb{P}(y^{n-1}) \\
&\qquad\qquad + D_{KL}\left(\mathbb{P}(y^{n-1}), \mathbb{Q}(y^{n-1})\right) \\
&\vdots
\end{aligned}
$$

□

# (3) Chain rule for KL-divergence

Here,

$$
\begin{aligned}
&\sum_{t=1}^{n} \mathbb{P} D_{KL}\left(\mathbb{P}(Y_t|Y^{t-1}), \mathbb{P}_{j^*}(Y_t|Y^{t-1})\right) \\
&= \sum_{t=1}^{n} \left( 0 \times \mathbb{P}\{I_t \neq j^*\} + d\left(\frac{1}{2}, \frac{1}{2} + \epsilon\right) \mathbb{P}\{I_t = j^*\} \right) \\
&= \sum_{t=1}^{n} \mathbb{P}\{I_t = j^*\} \frac{1}{2} \log \frac{1}{1 - 4\epsilon^2} \\
&= \frac{1}{2} \log\left(\frac{1}{1 - 4\epsilon^2}\right) \mathbb{P} T_{j^*}(n).
\end{aligned}
$$