Correction of the three exercises around the upper bound

 Lecture 1 — Winter 2022 term Lecture 1 — Winter 2022 term

Exercise #1 / Various considerations around the notion of regret.

1) ($N = 2$ is enough)

Time	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	etc.
ℓ_{1t}	1	0	1	0	1	0	
$\sum_{s=1}^{t-1} \ell_{1s}$		1	1	2	2	3	
ℓ_{2t}	$1/2$	1	0	1	0	1	
$\sum_{s=1}^{t-1} \ell_{2s}$		$1/2$	$1+1/2$	$1+1/2$	$2+1/2$	$2+1/2$	
Leader over $s=1,\ldots,t-1$		2	1	2	1	2	
p_{1t}	$1/2$	0	1	0	1	0	
p_{2t}	$1/2$	1	0	1	0	1	
$\sum_{j=1}^{2} p_{jt}\,\ell_{jt}$	$\frac{3}{4}$	1	1	1	1	1	
$\sum_{s=1}^{t} \sum_{j} p_{jt}\,\ell_{jt}$	$\frac{3}{4}$	$1+\frac{3}{4}$	$2+\frac{3}{4}$	$3+\frac{3}{4}$	$4+\frac{3}{4}$	$5+\frac{3}{4}$	

Follow the leader is a strategy that fails miserably $\longrightarrow$ EWA is a smoothed alternative.

We have
$$\sum_{t=1}^{T} \sum_{j=1}^{2} p_{jt}\,\ell_{jt} = T-1 + \frac{3}{4}$$

$$\sum_{t=1}^{T} \ell_{1t} = \begin{cases} T/2 & \text{if } T \text{ even} \\ \frac{T-1}{2} & \text{if } T \text{ odd} \end{cases} \qquad \text{(lower integer part)}$$

$$\sum_{t=1}^{T} \ell_{2t} = \begin{cases} (T-1)/2 + 1/2 & \text{if } T \text{ odd} \\ T/2 - 1 + 1/2 & \text{if } T \text{ even} \end{cases}$$

$$\min\left\{ \sum_{t=1}^{T} \ell_{1t},\ \sum_{t=1}^{T} \ell_{2t} \right\} \leqslant \frac{T}{2}$$

$$R_T \geqslant T-1+\frac{3}{4} - T/2 = T/2 - 1/4 \qquad \neq o(T)$$

2) It suffices to consider only the sequences in $\{0,1\}^{NT}$:
For any given strategy, we denote for $t \geqslant 2$:

$$k_t^* \in \operatorname*{argmax}_{k \in \{1..N\}} p_{kt} \longrightarrow \left[\begin{array}{l}\text{in particular,} \\ p_{k_t^*,t} \geqslant 1/N\end{array}\right.$$

Note that k_t^* depends only on ℓ_{js}, $\begin{array}{l} j \in \{1..N\} \\ s \in \{1..t-1\}\end{array}$

The sequence $\left\{\begin{array}{ll} \ell_{k_t^*,t} = 1 & \\ \ell_{jt} = 0 & j \neq k_t^* \end{array}\right.$

is such that: $\displaystyle \sum_{t=1}^{T} \sum_j p_{jt}\, \ell_{jt} \geqslant \sum_{t=1}^{T} \underbrace{p_{k_t^*,t}}_{\geqslant 1/N} \ell_{k_t^*,t} \geqslant \frac{T}{N}$

while $\displaystyle \sum_{t=1}^{T} \min_j \ell_{jt} = 0.$

Thus, no strategy can be such that

$$\sup_{\substack{(\ell_{1t} .. \ell_{Nt}) \in \{0,1\}^N \\ t=1..T}} \left\{ \sum_{t,j} p_{jt}\, \ell_{jt} - \sum_t \min_k \ell_{kt} \right\} = o(T)$$

Lecture 1 — Winter 2022 term Lecture 1 — Winter 2022 term

3) In the proof, instead of applying Hoeffding's lemma

$$\ln \mathbb{E}\left[e^{\eta X}\right] \leqslant \eta\, \mathbb{E}X + \frac{\eta^2}{8}(M-m)^2$$

we apply Jensen's inequality: $\ln \mathbb{E}\left[e^{\eta X}\right] \geqslant \eta\, \mathbb{E}X$

(valid $\forall \eta \in \mathbb{R}$ and all variables X s.t. X is integrable)

Then
$$\sum_j p_{j,t}\,\ell_{j,t} = -\frac{1}{\eta}\underbrace{\left(-\eta \sum_j p_{j,t}\,\ell_{j,t}\right)}_{\leqslant \ln \sum_j p_{j,t} e^{-\eta \ell_{j,t}}} \geqslant -\frac{1}{\eta} \ln \sum_j p_{j,t} e^{-\eta \ell_{j,t}}$$

Now, with the same telescoping argument:

$$\sum_{t=1}^{T}\sum_{j=1}^{N} p_{j,t}\,\ell_{j,t} \geqslant -\frac{1}{\eta} \ln \frac{\sum_{j=1}^{N} e^{-\eta \sum_{t=1}^{T} \ell_{j,t}}}{N}$$

$$\geqslant -\frac{1}{\eta} \ln \max_{j=1\ldots N} e^{-\eta \sum_{t=1}^{T} \ell_{j,t}}$$

upper bound the average in the log by the max

$$= \min_{j=1\ldots N} \sum_{t=1}^{T} \ell_{j,t}$$

That, $\forall \eta > 0$, $R_T = \sum_{t,j} p_{j,t}\,\ell_{j,t} - \min_k \sum_{t=1}^{T} \ell_{k,t} \geqslant 0$

 Lecture 1 — Winter 2022 term Lecture 1 — Winter 2022 term

Correction of the non-asymptotic lower bound

(1) $\sup_{\ell_{jt} \in [0,1]} \{ \dots \} \geqslant \mathbb{E}[\quad]$ for any probability distribution over the losses

Thus, $\sup_{\ell_{jt} \in [0,1]} \left\{ \sum_{t,j} p_{jt}\ell_{jt} - \min_i \sum \ell_{it} \right\}$

$$(*) \qquad \geqslant \max_{k \in \{1,\dots,N\}} \mathbb{E}_k \left[\sum_{t,j} p_{jt} L_{jt} - \min_i \sum_{t=1}^{T} L_{it} \right]$$

Denote by $\mathcal{F}_{t-1} = \sigma\left(L_{js},\ j \in \{1..N\} \text{ and } s \in \{1,\dots,t-1\} \right)$ for $t \geqslant 2$

For $t \geqslant 2$: p_t is $\mathcal{F}_{t-1}$-measurable, so that by the tower rule:

$$\mathbb{E}_k\left[\sum_{j=1}^{N} p_{jt} L_{jt} \right] = \mathbb{E}_k\left[\mathbb{E}\left[\sum_j p_{jt} L_{jt} \,\middle|\, \mathcal{F}_{t-1} \right] \right]$$

$$= \mathbb{E}_k\left[\sum_j p_{jt} \underbrace{\mathbb{E}_k\left[L_{jt} \,\middle|\, \mathcal{F}_{t-1} \right]} \right]$$

$= \mathbb{E}_k[L_{jt}]$ by independence of the losses across time

$$= \begin{cases} 1/2 & \text{if } j \neq k \\ 1/2 - \varepsilon & \text{if } j = k \end{cases}$$

Thus, $$\mathbb{E}_k\left[\sum_{j=1}^{N} p_{jt} L_{jt} \right] = \mathbb{E}_k\left[\sum_{j \neq k} p_{jt}/2 + p_{kt}\left(1/2 - \varepsilon\right) \right]$$

$$= 1/2 - \varepsilon\, \mathbb{E}_k[p_{kt}] \qquad (**)$$

This is also ok for $t=1$ (in that case, p_1 is constant).

On the other hand, $\mathbb{E}_k\left[\min_{i \in \{1..N\}} \sum_{t=1}^{T} L_{it} \right]$

$$\leqslant \min_{i \leq N} \underbrace{\mathbb{E}_k\left[\sum_{t=1}^{T} L_{it} \right]} = T/2 - \varepsilon T \qquad (\overset{*}{**})$$

$$= \begin{cases} T/2 & \text{if } i \neq k \\ T/2 - \varepsilon T & \text{if } i = k \end{cases}$$

Substituting $(**)$ and $(\overset{*}{**})$ in $(*)$, we get:

$$\sup_{\ell_{jt}\in[0,1]} \left\{ \sum_{t,j} p_{jt}\,\ell_{jt} - \min_i \sum_{t=1}^{T} \ell_{it} \right\}$$

$$\geqslant \max_{k\leq N} \mathbb{E}_k \left[\sum_{t,j} p_{jt} L_{jt} - \min_{i\leq N} \sum_{t=1}^{T} L_{it} \right]$$

$$\geqslant \max_{k\leq N} T\varepsilon \left(1 - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_k [p_{kt}] \right)$$

don't forget that for $t \geqslant 2$, p_{kt} is a random variable, as it depends on the L_{js}, $j \leq N$ and $s \leq t-1$

IDEA: The idea of the proof is that any strategy will take some time (basically, a time of order $\sqrt{T}$) to identify k as the best arm in $\{1, \dots N\}$ under $\mathbb{P}_k$.

Since this needs to be performed for N distributions $\mathbb{P}_1, \dots \mathbb{P}_N$ at a time, an additional $\sqrt{\ln N}$ factor will be gained by Fano's lemma.

(2) Deux choses à voir :

- $\mathrm{kl}(p,q) \leqslant \overline{K}$
- $a \leqslant \mathrm{kl}(p,q)/\ln N$ lorsque $a \geqslant 2e/(2e+1)$

2nd point :

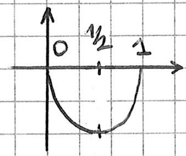
$$\mathrm{kl}(p,q) = p \ln p + (1-p)\ln(1-p) + p \ln \frac{1}{q} + (1-p) \ln\left(\frac{1}{1-q}\right)$$

avec $p = \frac{1}{N-1} \sum_{j\geqslant 2} Q_j(A_j) \geqslant a = \min_{j=1,\dots N} Q_j(A_j)$

et $q = \frac{1}{N-1} \sum_{j\geqslant 2} Q_1(A_j) = \frac{1}{N-1} \left(1 - Q_1(A_1) \right) \leqslant \frac{1}{N-1} (1-a)$

cf. (A_k) est une partition de Ω

We use that $x \mapsto x \ln x + (1-x) \ln x$ is increasing on $[1/2, 1]$



to get $p \ln p + (1-p) \ln(1-p) \geqslant a \ln a + (1-a) \ln(1-a)$

We use $(1-p) \ln \dfrac{1}{1-q} \geqslant 0$

We have $p \geqslant a$ and $\dfrac{1}{q} \geqslant \dfrac{N-1}{1-a} \geqslant 1$ so that $p \ln \dfrac{1}{q}$

$\geqslant a \ln\left(\dfrac{N-1}{1-a}\right)$

All in all, $\mathrm{kl}(p,q) \geqslant a \ln a + (1-a) \ln(1-a) + a \ln\left(\dfrac{N-1}{1-a}\right)$

$N-1 \geqslant \dfrac{N}{2} \quad \forall N \geqslant 2$

and $a \ln(N-1) \geqslant a \ln N - a \ln 2$

$$\geqslant a \ln N + \Big[\underbrace{a \ln a + (1-a)\ln(1-a) - a \ln(1-a) - a \ln 2}_{\text{function study: this is } \geqslant 0 \text{ for } a \geqslant 0.76 \text{ (while } 2e/(2e+1) \simeq 0.845)} \Big]$$

or (original Birgé's argument)

$(1-a)\ln(1-a) \geqslant \min\limits_{t \in [0,1]} t \ln t = -1/e$

$\geqslant -2e/(2e+1) \geqslant -a = a \ln \dfrac{1}{e}$

thus

$a \ln a + (1-a)\ln(1-a) - a \ln(1-a) - a \ln 2$

$\geqslant a \ln \dfrac{a}{2e(1-a)} \geqslant a \ln\left(\dfrac{1}{2e} \dfrac{2e/(2e+1)}{1 - 2e/(2e+1)}\right) = 0$

$t \mapsto \dfrac{t}{1-t}$ increasing and $a \geqslant 2e/2e+1$

In any case:

$$\mathrm{kl}(p,q) \geqslant a \ln N \quad \text{when (eg)} \quad a \geqslant \frac{2e}{2e+1}.$$

1st point $\hookrightarrow$ We now prove that $\mathrm{kl}(p,q) \leqslant \overline{K}$.

The data compression inequality entails that:

- for all distributions μ, ν on $(\Omega, \mathcal{F})$ and any $A \in \mathcal{F}$,

$$\mathrm{kl}\big(\mu(A), \nu(A)\big) \leq \mathrm{KL}(\mu, \nu)$$

indeed, consider $X = \mathbb{1}_A$, then $\mu^X = \mu^{\mathbb{1}_A}$ is the Bernoulli distribution with parameter $\mu(A)$; same for ν; thus:

$$\mathrm{kl}\big(\mu(A), \nu(A)\big) \underset{\substack{\uparrow \\ \text{by definition}}}{=} \mathrm{KL}\big(\mu^{\mathbb{1}_A}, \nu^{\mathbb{1}_A}\big) \underset{\substack{\uparrow \\ \text{data-} \\ \text{compression inequality}}}{\leq} \mathrm{KL}(\mu, \nu)$$

- KL (and thus kl) is jointly convex:

$\forall \alpha \in (0,1)$, $\forall \mu_1, \mu_2$, $\forall \nu_1, \nu_2$,

$$\mathrm{KL}\big(\alpha\mu_1 + (1-\alpha)\mu_2, \ \alpha\nu_1 + (1-\alpha)\nu_2\big) \leq \alpha \, \mathrm{KL}(\mu_1, \nu_1) + (1-\alpha)\, \mathrm{KL}(\mu_2, \nu_2) \qquad (C)$$

other, more direct, proofs exist.

Proof: $\Omega' = \Omega \times \{1,2\}$

$\tilde{\mu}$ on Ω' given by $\forall A \in \mathcal{F}$: $\tilde{\mu}\big(A \times \{j\}\big) = \begin{cases} \alpha\mu_1(A) & \text{if } j=1 \\ (1-\alpha)\mu_2(A) & \text{if } j=2 \end{cases}$

same for $\tilde{\nu}$ based on ν

Let π be the projection $(\omega, j) \in \Omega \times \{1,2\} \longmapsto \omega$

Then $\tilde{\mu}^\pi$ = 1^{st} marginal of $\tilde{\mu}$ = $\alpha\mu_1 + (1-\alpha)\mu_2$

$\tilde{\nu}^\pi$ = $\tilde{\nu}$ = $\alpha\nu_1 + (1-\alpha)\nu_2$

The desired inegality holds by data compression:

$$\mathrm{KL}\big(\tilde{\mu}^\pi, \tilde{\nu}^\pi\big) = \mathrm{KL}\big(\alpha\mu_1 + (1-\alpha)\mu_2, \ \alpha\nu_1 + (1-\alpha)\nu_2\big)$$
$$\leq \mathrm{KL}(\tilde{\mu}, \tilde{\nu}) = ?$$

With no loss of generality we can assume $\mu_1 \ll \nu_1$ and $\mu_2 \ll \nu_2$ otherwise the desired inequality (C) is satisfied (its right-hand side $= +\infty$)

Then $\tilde{\mu} \ll \tilde{\nu}$ as well, with $\dfrac{d\tilde{\mu}}{d\tilde{\nu}}(\omega, j) = \dfrac{d\mu_j}{d\nu_j}(\omega)$

$$\mathrm{KL}(\tilde{\mu}, \tilde{\nu}) = \int_{\Omega \times \{1,2\}} \left(\ln \frac{d\tilde{\mu}}{d\tilde{\nu}} \right) d\tilde{\nu}$$

$$= \alpha \times \int_{\Omega} \left(\ln \frac{d\mu_1}{d\tilde{\nu}_1} \right) d\tilde{\nu}_1 + (1-\alpha) \int_{\Omega} \left(\ln \frac{d\mu_2}{d\tilde{\nu}_2} \right) d\tilde{\nu}_2$$

$$= \alpha \; \mathrm{KL}(\mu_1, \tilde{\nu}_1) + (1-\alpha)\, \mathrm{KL}(\mu_2, \tilde{\nu}_2).$$

Application: $\mathrm{kl}(p, q) = \mathrm{kl}\left(\frac{1}{N-1} \sum_{j \geq 2} Q_j(A_j), \frac{1}{N-1} \sum_{j \geq 2} Q_1(A_j) \right)$

$$\underset{\substack{\text{joint}\\ \text{convexity}\\ \text{of kl}}}{\leq} \frac{1}{N-1} \sum_{j \geq 2} \mathrm{kl}\big(Q_j(A_j), Q_1(A_j)\big)$$

$$\underset{\substack{\text{data-}\\ \text{compression ineq.}}}{\leq} \frac{1}{N-1} \sum_{j \geq 2} \mathrm{KL}(Q_j, Q_1) = \overline{K}.$$

(3) We denote $a_k = \mathbb{E}_k\left[\frac{1}{T} \sum_{t=1}^{T} p_{kt} \right]$ and $b_k = \mathbb{E}_1\left[\frac{1}{T} \sum_{t=1}^{T} p_{kt} \right]$
$(k \geq 1)$ note that $b_1 = a_1$

The "2nd point" of the proof in (2) was purely analytical and only used that $b_1 + b_2 + \cdots + b_N = 1$, which is still true.

Therefore, we similarly get:

$$\min_{k=1,\dots,N} \mathbb{E}_k\left[\frac{1}{T} \sum_{t=1}^{T} p_{kt} \right] \leq \max\left\{ \frac{2e}{2e+1}, \mathrm{kl}\left(\frac{1}{N-1} \sum_{j \geq 2} \mathbb{E}_j\left[\frac{1}{T} \sum_t p_{jt} \right], \frac{1}{N-1} \sum_{j \geq 2} \mathbb{E}_1\left[\frac{1}{T} \sum_t p_{jt} \right] \right) \times 1/\ln N \right\}$$

By convexity of kl, we may further upper bound the right-hand side by

$$\max\left\{ \frac{2e}{2e+1}, \frac{1}{\ln N} \left(\frac{1}{N-1} \sum_{j=2}^{N} \mathrm{kl}\left(\mathbb{E}_j\left[\frac{1}{T} \sum_t p_{jt} \right], \mathbb{E}_1\left[\frac{1}{T} \sum_t p_{jt} \right] \right) \right) \right\}$$

It thus suffices to show that $\mathrm{kl}(\mathbb{E}_j Z, \mathbb{E}_1 Z) \leq \mathrm{KL}(\mathbb{P}_j^L, \mathbb{P}_1^L)$ for any random variable Z that

- takes values in $[0,1]$
- is $\sigma(L)$-measurable

that is,

 Lecture 1 — Winter 2022 term Lecture 1 — Winter 2022 term

random variables Z of the form $Z = \Psi(L)$
where $\Psi : \ell = (\ell_{j,t})_{\substack{j \leq N \\ t \leq T}} \longmapsto \Psi(\ell) \in [0,1]$ is measurable.

But $\mathbb{E}_j Z = \mathbb{E}_j \Psi(L)$

$$= \int \Psi(\ell) \, d\mathbb{P}_j^L(\ell)$$

where $\mathbb{P}_j^L$ is the image distribution of $\mathbb{P}_j$ by L.

The result thus follows from the two reminders of the property of the KL divergence (see first page of the statement of the exercise):

<u>Lemma</u>: Let $f : (\Omega, \mathcal{F}) \longrightarrow [0,1]$ be measurable and let μ, ν be probability distributions over $(\Omega, \mathcal{F})$. Then:

$$\mathrm{kl}\left(\int_\Omega f \, d\mu, \int_\Omega f \, d\nu \right) \leq \mathrm{KL}(\mu, \nu).$$

<u>Proof</u>: Let $\tilde{\Omega} = \Omega \times [0,1]$ (equipped with the product σ-algebra)

Let $E = \{ (\omega, t) \text{ s.t. } f(\omega) \geq t \}$; E is measurable

Let $\tilde{\mu} = \mu \otimes \lambda$ and $\tilde{\nu} = \nu \otimes \lambda$ where λ is the Lebesgue measure on $[0,1]$

$$\mathrm{KL}(\mu, \nu) = \mathrm{KL}(\tilde{\mu}, \tilde{\nu}) \geq \mathrm{kl}\big(\tilde{\mu}(E), \tilde{\nu}(E)\big)$$

cf. 1st reminder on KL divergence as $\mathrm{KL}(\lambda, \lambda) = 0$

a special case of data compression with $X = \mathbb{1}_E$

But by Fubini–Tonelli :
$$\tilde{\mu}(E) = \iint \mathbb{1}_{\{f(\omega) \geq t\}} \, d\mu(\omega) \, dt = \int f(\omega) \, d\mu(\omega)$$
integrating over t

and similarly for $\tilde{\nu}(E)$.

(4) By independence, $\mathbb{P}_j^L$ is a product of NT distributions. Using that $KL(\mu \otimes \mu', \nu \otimes \nu') = KL(\mu,\nu) + KL(\mu', \nu')$ (iterating this equality), we get

$$KL(\mathbb{P}_j^L, \mathbb{P}_1^L) = \sum_{k,t} \, KL(\mathbb{P}_j^{L_{kt}}, \mathbb{P}_1^{L_{kt}})$$

$$\uparrow \begin{cases} = 0 & \text{if } k \neq 1 \text{ and } k \neq j \\ = KL(\mathrm{Ber}(1/2-\varepsilon), \mathrm{Ber}(1/2)) & \text{if } k = j \\ = KL(\mathrm{Ber}(1/2), \mathrm{Ber}(1/2-\varepsilon)) & \text{if } k = 1 \end{cases}$$

Thus, $\forall j$, $KL(\mathbb{P}_j^L, \mathbb{P}_1^L) = T \times \Big(\underbrace{kl(1/2-\varepsilon, 1/2) + kl(1/2, 1/2-\varepsilon)} \Big)$

it suffices to show that this is $\leqslant 5\varepsilon^2$ when $\varepsilon \leqslant 1/10$.

$$kl(1/2-\varepsilon, 1/2) + kl(1/2, 1/2-\varepsilon)$$

$$= \left(\frac{1}{2}-\varepsilon\right) \ln \frac{1/2-\varepsilon}{1/2} + \left(1-(1/2-\varepsilon)\right) \ln \frac{1-(1/2-\varepsilon)}{1/2} + \frac{1}{2} \ln \frac{1/2}{1/2-\varepsilon} + \frac{1}{2} \ln \frac{1/2}{1-(1/2-\varepsilon)}$$

$$= (1/2-\varepsilon) \ln(1-2\varepsilon) + (1/2+\varepsilon) \ln(1+2\varepsilon) - \frac{1}{2}\ln(1-2\varepsilon) - \frac{1}{2}\ln(1+2\varepsilon)$$

$$= \varepsilon \ln(1+2\varepsilon) - \varepsilon \ln(1-2\varepsilon) = \varepsilon \ln\big((1+2\varepsilon)/(1-2\varepsilon) \big)$$

$$= \varepsilon \ln\left(1 + \frac{4\varepsilon}{1-2\varepsilon}\right) \underset{\ln(1+u) \leqslant u}{\leqslant} \frac{4\varepsilon^2}{1-2\varepsilon} \leqslant 5\varepsilon^2$$

$\ln(1+u) \leqslant u$; for $\varepsilon \leqslant 1/10$, $1-2\varepsilon \geqslant 4/5$

Hence $\overline{K}' = \dfrac{1}{N-1} \displaystyle\sum_{j \geqslant 2} KL(\mathbb{P}_j^L, \mathbb{P}_1^L)$

$$= T\big(kl(1/2-\varepsilon, 1/2) + kl(1/2, 1/2-\varepsilon) \big) \leqslant 5T\varepsilon^2 \quad \text{for } \varepsilon \leqslant 1/10.$$

(5) Questions (1)–(4) lead to $\forall \varepsilon \in (0, 1/10]$,

$$SR_T \overset{\text{def}}{=} \sup_{l_{jt} \in [0,1]} \left\{ \sum_{t,j} p_{jt}\, l_{jt} - \min_k \sum_{t=1}^{T} l_{kt} \right\} \geqslant T\varepsilon \left(1 - \min_k \mathbb{E}_k\left[\frac{1}{T} \sum_t p_{kt} \right] \right)$$

$$\geqslant T\varepsilon \left(1 - \max\left\{ \frac{2e}{2e+1}, \frac{5T\varepsilon^2}{\ln N} \right\} \right)$$

 Lecture 1 — Winter 2022 term Lecture 1 — Winter 2022 term

We would like to take, e.g., ε such that $5T\varepsilon^2 / \ln N = \dfrac{2e}{2e+1}$

that is, $\varepsilon^* = \sqrt{\dfrac{2e}{2e+1} \dfrac{\ln N}{5T}}$

This ε^* is $\leq 1/10$ when

$$\frac{\ln N}{T} \leq \frac{(2e+1)5}{2e} \times \frac{1}{100} \approx 0{,}059197$$

and $1/17 \approx 0{,}58823$

Thus, $\varepsilon^* \leq 1/10$ when $T \geq 17 \ln N$.

With this ε^*, the bound becoms $T\,\varepsilon^*\left(1 - \dfrac{2e}{2e+1}\right)$

$$= \sqrt{T \ln N} \times \underbrace{\left(\sqrt{\frac{2e}{(2e+1)5}} \times \frac{1}{2e+1} \right)}_{\geq 0.06}$$

<u>Theorem</u> For all strategis, for all $N \geq 2$, for all $T \geq 17 \ln N$,

$$\sup_{\ell_{jt} \in [0,1]} \left\{ \sum_{t,j} p_{jt}\, \ell_{jt} - \min_k \sum_t \ell_{kt} \right\} \geq 0.06 \sqrt{T \ln N}$$

<u>PS</u> There will be bonus points for those who will significantly improve both constants 17 and 0.06! In particular, the 0.06 should become as close as possible to $1/\sqrt{2} \simeq 0{,}7$.