Adversarial Online Learning with noise

Supplementary Manuscript

Proof of Lemma 3

The proof follows the standard analysis of exponential weighting schemes: let $W_t = \sum_{i=1}^K w_{i,t}$ using the algorithm update we can write

$$\begin{aligned}
\frac{W_{t+1}}{W_t} &= \sum_{i=1}^K \frac{w_{i,t+1}}{W_t} = \sum_{i=1}^K \frac{w_{i,t} e^{-\eta \hat{\ell}_{i,t}}}{W_t} = \sum_{i=1}^K q_{i,t} e^{-\eta \hat{\ell}_{i,t}} \\
&\leq \sum_{i=1}^K q_{i,t}(1 - \eta \hat{\ell}_{i,t} + \eta^2 (\hat{\ell}_{i,t})^2) \qquad (\text{using } e^x \leq 1 + x + x^2 \text{ for } x \leq 1) \\
&= 1 - \eta \sum_{i=1}^K q_{i,t} \hat{\ell}_{i,t} + \eta^2 \sum_{i=1}^K q_{i,t} (\hat{\ell}_{i,t})^2
\end{aligned}$$

Taking logs and using $\ln(1-x) \leq -x$ for all x and summing for $t = 1, 2, ..., T$ yields

$$\ln \frac{W_{T+1}}{W_1} \leq -\eta \sum_{t=1}^T \sum_{i=1}^K q_{i,t} \hat{\ell}_{i,t} + \eta^2 \sum_{t=1}^T \sum_{i=1}^K q_{i,t} (\hat{\ell}_{i,t})^2$$

Moreover, for any fixed action k we have $W_t \geq w_{k,t}$, thus:

$$\ln \frac{W_{T+1}}{W_1} \geq \ln \frac{w_{k,T+1}}{W_1} = -\eta \sum_{t=1}^T \hat{\ell}_{k,t} - \ln K$$

Putting together and rearranging gives:

$$\sum_{t=1}^T \sum_{i=1}^K q_{i,t} \hat{\ell}_{i,t} - \sum_{t=1}^T \hat{\ell}_{k,t} \leq \frac{\ln K}{\eta} + \eta \sum_{t=1}^T \sum_{i=1}^K q_{i,t} (\hat{\ell}_{i,t})^2$$

■

Proof of Theorem 2

Since

$$\hat{\ell}_{i,t} = \frac{c_{i,t} - p}{1 - 2p} \in \{\frac{1-p}{1-2p}, \frac{-p}{1-2p}\} = \{-\frac{1-\epsilon}{2\epsilon}, \frac{1+\epsilon}{2\epsilon}\},$$

we have that

$$-\eta \hat{\ell}_{i,t} \leq \epsilon \sqrt{\frac{\ln K}{T}} \frac{1-\epsilon}{2\epsilon} \leq \frac{\sqrt{\frac{\ln K}{T}}}{2} \leq 1$$

where the last equation uses $T \geq \frac{1}{4} \ln K$. Thus, we can apply Lemma 3 and obtain

$$\sum_{t=1}^T \sum_{i=1}^K q_{i,t} \hat{\ell}_{i,t} - \sum_{t=1}^T \hat{\ell}_{k,t} \leq \frac{\ln K}{\eta} + \eta \sum_{t=1}^T \sum_{i=1}^K q_{i,t} (\hat{\ell}_{i,t})^2$$

Taking expectation on both sides and using that the estimator is unbiased (i.e., $\mathrm{E}[\hat{\ell}_{i,t}] = \ell_{i,t}$) yields,

$$\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\ell_{i,t} - \sum_{t=1}^{T}\ell_{k,t} \leq \frac{\ln K}{\eta} + \eta\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[(\hat{\ell}_{i,t})^2]$$

Using the fact that $Regret(T) = \sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\ell_{i,t} - \min_{k\in A}\sum_{t=1}^{T}\ell_{k,t}$, we have,

$$Regret(T) \leq \frac{\ln K}{\eta} + \eta\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[(\hat{\ell}_{i,t})^2]$$

We bound the second moments of the estimate as follows,

$$\mathrm{E}[(\hat{\ell}_{i,t})^2] = p\frac{(\bar{\ell}_{i,t}-p)^2}{(1-2p)^2} + (1-p)\frac{(\ell_{i,t}-p)^2}{(1-2p)^2} \leq \frac{1}{(1-2p)^2} = \frac{1}{\epsilon^2}\ ,$$

where $\bar{\ell}_{i,t} = 1 - \ell_{i,t}$. Putting it back together and plugging $\eta = \epsilon\sqrt{\frac{\ln K}{T}}$ we obtain

$$Regret(T) \leq \frac{\ln K}{\eta} + \frac{\eta T}{\epsilon^2} \leq \frac{2}{\epsilon}\sqrt{T\ln K}$$

■

Proof of Theorem 4

By applying Lemma 3 and taking expectation on both sides we obtain

$$\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[\hat{\ell}_{i,t}] - \sum_{t=1}^{T}\mathrm{E}[\hat{\ell}_{k,t}] \leq \frac{\ln K}{\eta} + \eta\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[(\hat{\ell}_{i,t})^2]$$

Calculating the expectation of the estimator $\hat{\ell}_{i,t}$, and since $\ell_{i,t} \in \{0,1\}$, we have,

$$\mathrm{E}[\hat{\ell}_{i,t}] = (1-p)\ell_{i,t} + p\bar{\ell}_{i,t} = (1-2p)\ell_{i,t} + p = |\ell_{i,t} - p|$$

For the second moment we have $(\hat{\ell}_{i,t})^2 = c_{i,t}^2 = c_{i,t} \leq 1$. Putting things together we have

$$\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}|\ell_{i,t}-p| - \sum_{t=1}^{T}|\ell_{k,t}-p| \leq \frac{\ln K}{\eta} + \eta T \tag{1}$$

Using the notation of $\hat{L}_{ON,T} = \sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\hat{\ell}_{i,t}$, $\hat{L}_{k,T} = \sum_{t=1}^{T}\hat{\ell}_{k,t}$, $L_{ON,T} = \sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\ell_{i,t}$, and $L_{k,T} = \sum_{t=1}^{T}\ell_{k,t}$, we can write inequality (1) as

$$\mathrm{E}[\hat{L}_{ON,T}] - \mathrm{E}[\hat{L}_{k,T}] \leq \frac{\ln K}{\eta} + \eta T$$

Denote by $G_{t,b} = \{i \in A \mid \ell_{i,t} = b\}$ the set of actions with loss $b \in \{0,1\}$ in round t. Denote by $Q_t = \sum_{i\in G_{t,1}} q_{i,t}$ the distribution mass the learner gives actions in $G_{t,1}$. Using

this notation we have $L_{ON,T} = \sum_{t=1}^{T} Q_t$. Now calculate the value of the estimated losses of the online algorithm,

$$\begin{aligned}
\mathrm{E}[\hat{L}_{ON,T}] &= \sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}|\ell_{i,t} - p| = \sum_{t=1}^{T}[p \sum_{i\in G_{t,0}} q_{i,t} + (1-p)\sum_{i\in G_{t,1}} q_{i,t}] \\
&= \sum_{t=1}^{T}[p(1-Q_t) + (1-p)Q_t] = \sum_{t=1}^{T}[p + (1-2p)Q_t] \\
&= (1-2p)L_{ON,T} + pT
\end{aligned}$$

Similarly, for the term $\mathrm{E}[\hat{L}_{k,T}]$ we have,

$$\begin{aligned}
\mathrm{E}[\hat{L}_{k,T}] &= \sum_{t=1}^{T} |\ell_{k,t} - p| = \sum_{t|\ell_{t,k}=0} p + \sum_{t|\ell_{t,k}=1} (1-p) \\
&= p(T - L_{k,T}) + (1-p)L_{k,T} = (1-2p)L_{k,T} + pT
\end{aligned}$$

Putting all together,

$$\mathrm{E}[\hat{L}_{ON,T}] - \mathrm{E}[\hat{L}_{k,T}] = (1-2p)L_{ON,T} + pT - [(1-2p)L_{k,T} + pT] = (1-2p)[L_{ON,T} - L_{k,T}]$$

Dividing by both sides of inequity by $(1-2p)$ and using $\eta = \sqrt{\frac{\ln K}{T}}$ we obtain that

$$Regret(T) = L_{ON,T} - \min_{k\in A} L_{k,T} \leq \frac{1}{1-2p}(\frac{\ln K}{\eta} + \eta T) = \frac{2}{\epsilon}\sqrt{T \ln K}$$

■

Proof of Theorem 5

We first prove for $K \geq 27$. To prove the theorem we first define the following adversarial loss assignment strategy:

- the adversary initially picks uniformly a *best action* $i^\star$ ($\forall i\ Pr[i^\star = i] = \frac{1}{K}$)
- at round t: the adversary draws losses for the actions from the following distributions:
  1. for $i^\star$: $\ell_{i^\star,t} \sim B(\frac{1}{2} - \delta)$
  2. for $i \neq i^\star$: $\ell_{i,t} \sim B(\frac{1}{2})$

where $\delta = \min\{\frac{1}{6\epsilon}\sqrt{\frac{\ln K}{T}}, \frac{1}{2}\}$. Now we calculate the distribution of the **ϵ-noisy feedback** $c_{i,t}$. Starting with the best action we have

$$\begin{aligned}
Pr[c_{i^\star,t} = 1] &= Pr[\ell_{i^\star,t} = 1]Pr[R_\epsilon = 0] + Pr[\ell_{i^\star,t} = 0]Pr[R_\epsilon = 1] \\
&= (\frac{1}{2} - \delta)\frac{1+\epsilon}{2} + (\frac{1}{2} + \delta)\frac{1-\epsilon}{2} = \frac{1}{2} - \epsilon\delta
\end{aligned}$$

For $i \neq i^\star$ we have

$$\begin{aligned} Pr[c_{i,t}=1] &= Pr[\ell_{i,t}=1]Pr[R_\epsilon=0] + Pr[\ell_{i,t}=0]Pr[R_\epsilon=1] \\ &= \frac{1}{2}\frac{1+\epsilon}{2} + \frac{1}{2}\frac{1-\epsilon}{2} = \frac{1}{2} \end{aligned}$$

Thus, we have: $c_{i^\star,t} \sim B(\frac{1}{2}-\epsilon\delta)$ and $c_{i,t} \sim B(\frac{1}{2})$ for $i \neq i^\star$.

The following is a standard claim regarding the minimum of i.i.d binomial random variables.

Lemma 1 *Let $X_1,...,X_{K-1}$ be i.i.d random variables with distribution $B(n,p)$ such that $p \in (\frac{1}{4},\frac{1}{2})$, $n \geq 2\ln K$ and $K \geq e^{27}$. Then with probability of at least $\frac{1}{2}$ we have*

$$\min\{X_1,...,X_{K-1}\} \leq np - \sqrt{\frac{p}{9}n\ln K}$$

Proof Denote by $Y = \min\{X_1,...,X_{K-1}\}$ then by interdependency we can write

$$\begin{aligned} Pr[Y \leq np-t] &= 1 - Pr[\forall i \in \{1,2,...,K-1\} \;\; X_i \geq np-t] \\ &= 1 - (Pr[X_1 \geq np-t])^{K-1} \end{aligned} \tag{2}$$

Now we want to bound $Pr[X_1 \geq np-t]$. Rearranging, and using *Lemma 5.2* of Klein and Young (1999), for $t \leq \frac{1}{2}pn$ we can bound

$$\begin{aligned} Pr[X_1 \geq np-t] &= Pr[X_1 - np \geq -t] = 1 - Pr[X_1 - np \leq -t] \\ &\leq 1 - \exp(-\frac{9t^2}{np}) = 1 - \frac{1}{K} \end{aligned}$$

where in the last equation we take $t = \sqrt{\frac{p}{9}n\ln K} \leq \frac{1}{2}pn$. Plugging it back in (2) we obtain

$$Pr[Y \leq np - \sqrt{\frac{p}{9}n\ln K}] \geq 1 - (1-\frac{1}{K})^{K-1} \geq \frac{1}{2}$$

■

Denoting by $C_{i,T} = \sum_{t=1}^{T} c_{i,t}$, the sum of the noisy feedback of action i. Note that this is binomial random variable. In addition, for $i^\star$ we have $C_{i^\star,T} \sim B(T, \frac{1}{2}-\epsilon\delta)$ and for $i \neq i^\star$ we have $C_{i,T} \sim B(T,\frac{1}{2})$. By applying Lemma 1 on the noisy-feedbacks we show the following corollary.

Corollary 2 *With probability at least $\frac{1}{4}$ there exist action $j \neq i^\star$ such that $C_{j,T} < C_{i^\star,T}$.*

Proof Applying Lemma 1 on the $K-1$ actions with $c_{i,t} \sim B(\frac{1}{2})$ we obtain that with probability at least $\frac{1}{2}$ there exist action $j \neq i^\star$ such that

$$C_{j,T} \leq \frac{T}{2} - \sqrt{\frac{p}{9}T\ln K} < \frac{T}{2} - \frac{1}{6}\sqrt{T\ln K},$$

where the second inequality uses $p > 1/4$.
For the best action $i^\star$ we have $\mathrm{E}[C_{i^\star,T}] = \frac{T}{2} - \epsilon\delta T$.

- if $\delta = \frac{1}{6\epsilon}\sqrt{\frac{\ln K}{T}}$ we have

$$\mathrm{E}[C_{i^\star,T}] = \frac{T}{2} - \frac{1}{6}\sqrt{T \ln K}$$

Using the fact that for binomial distribution, $B(n,q)$, the median is $\lfloor nq \rfloor$ or $\lceil nq \rceil$ we have that with probability at least $\frac{1}{2}$

$$C_{i^\star,T} \geq \frac{T}{2} - \frac{1}{6}\sqrt{T \ln K}$$

- if $\delta = \frac{1}{2}$ we have that the distribution for the **ϵ-noisy feedback** of the *best action*, $c_{i^\star,t}$ is $B(\frac{1-\epsilon}{2})$, therefore

$$\mathrm{E}[C_{i^\star,T}] = \frac{T}{2} - \frac{\epsilon}{2}T$$

$\delta = \frac{1}{2}$ implies $\epsilon \leq \frac{1}{3}\sqrt{\frac{\ln K}{T}}$ (as $\delta = \min\{\frac{1}{6\epsilon}\sqrt{\frac{\ln K}{T}}, \frac{1}{2}\}$) thus,

$$\frac{\epsilon}{2}T \leq \frac{T}{2}\frac{1}{3}\sqrt{\frac{\ln K}{T}} = \frac{1}{6}\sqrt{T \ln K}$$

Therefore, we still have that with probability at least $\frac{1}{2}$

$$C_{i^\star,T} \geq \frac{T}{2} - \frac{1}{6}\sqrt{T \ln K}$$

Putting things together we obtain that with probability at least $\frac{1}{4}$ we have

$$C_{i^\star,T} > C_{j,T}$$

■

The following lemma states that the action that has smaller observed noisy-loss has a higher probability to be the *best action*.

Lemma 3 *Let $C_{1,T}, \ldots, C_{K,T}$ be a realization of the noisy-feedbacks, such that $C_{j_1,T} < C_{j_2,T}$, where $j_1, j_2 \in A$. Then,*

$$\Pr[i^\star = j_1 \mid C_{1,T}, \ldots, C_{K,T}] > \Pr[i^\star = j_2 \mid C_{1,T}, \ldots, C_{K,T}]$$

Proof Using Bayes' theorem we have for action $j \in A$ that

$$\begin{aligned}
\Pr[i^\star = j \mid C_{1,T}, \ldots, C_{K,T}] &= \frac{\Pr[C_{1,T}, \ldots, C_{K,T} \mid i^\star = j] \Pr[i^\star = j]}{\Pr[C_{1,T}, \ldots, C_{K,T}]} \\
&= \frac{\Pr[C_{j,T} \mid i^\star = j](\frac{1}{2}^T)^{K-1}\frac{1}{K}}{\Pr[C_{1,T}, \ldots, C_{K,T}]} = \frac{\Pr[C_{j,T} \mid i^\star = j]}{Z} \\
&= \frac{1}{Z}(\frac{1-\epsilon}{2})^{C_{j,T}}(\frac{1+\epsilon}{2})^{T-C_{j,T}}
\end{aligned}$$

where $Z = \frac{(\frac{1}{2}^T)^{K-1}\frac{1}{K}}{\Pr[C_{1,T},\ldots,C_{K,T}]}$ is a constant (not depend on j). Therefore, if $C_{j_1,T} < C_{j_2,T}$ then

$$\Pr[i^\star = j_1 \mid C_{1,T},\ldots,C_{K,T}] > \Pr[i^\star = j_2 \mid C_{1,T},\ldots,C_{K,T}]$$

■

Using the lemma we can show the following corollary.

Corollary 4 *consider an algorithm for "predicting the best action" problem: that is, the algorithm input is a realization $C_{1,T},\ldots,C_{K,T}$, i.e., for one action $i^\star$ we have $C_{i^\star,T} \sim B(T,\frac{1}{2}-\epsilon\delta)$ and for $j \neq i^\star$ we have $C_{j,T} \sim B(T,\frac{1}{2})$ and the output is an action I_T - a prediction for which action is optimal. Then for any algorithm we have,*

$$\Pr[I_T \neq i^\star] \geq \frac{1}{4}$$

where the probability is taken over the randomness of the algorithm, the losses, the noise and the draw of $i^\star$.

Proof Lemma 3 implies that the optimal algorithm will predict

$$I_T = \arg\min_{j\in A}\{C_{1,T},\ldots,C_{K,T}\}$$

From Corollary 2 we have that for the optimal algorithm

$$\Pr[I_T \neq i^\star] \geq \frac{1}{4}$$

■

Putting it all together we can now prove the theorem.

Proof of Theorem 5: For any round t we would think of the algorithm as algorithm for "predicting the best action" problem. Using this we can think of t as the the time horizon and by applying Corollary 4 conclude that for every t we have

$$\Pr[I_t \neq i^\star] \geq \frac{1}{4}$$

Therefore the expectation of the regret, when the expectation is taken over the losses, the noise and the draw of $i^\star$ (note that the regret itself includes the randomness of the algorithm) satisfies,

$$\mathrm{E}[Regret(T)] = \sum_{t=1}^{T}\Pr[I_t \neq i^\star]\delta \geq \frac{1}{4}T\delta,$$

where $\delta = \min\{\frac{1}{6\epsilon}\sqrt{\frac{\ln K}{T}},\frac{1}{2}\}$ concludes the proof. ■

Proof sketch for $K < 27$:
The proof is similar to the case of $K \geq 27$. For $2 \leq K < 27$ we need only to prove that

$$Regret(T) = \Omega(\min\{\frac{1}{\epsilon}\sqrt{T}, T\})$$

We use the same adversarial strategy as before, but with $\delta = \min\{\delta = \gamma\frac{1}{\epsilon}\sqrt{\frac{1}{T}}, \frac{1}{2}\}$ where $\gamma > 0$ is a constant. Taking any action i, such that $\ell_{i,t} \sim B(\frac{1}{2})$ (therefore also $c_{i,t} \sim B(\frac{1}{2})$), we have that $C_{i,T} \sim Bin(T, \frac{1}{2})$. We use the following lemma, which is a well-known fact about $Bin(T, \frac{1}{2})$ distribution.

Lemma 5 *Let $C \sim Bin(T, \frac{1}{2})$ then there exist a constant $\gamma > 0$ such that with probability $1/4$ we have*

$$C \leq \frac{T}{2} - \gamma\sqrt{T}$$

The lemma follows since for any k, we have that $\Pr[C = k] < \lambda/\sqrt{T}$, for some constant $\lambda > 0$.

Let γ be the constant of Lemma 5 and denote $\alpha = \Pr[C_{i,T} \leq \frac{T}{2} - \gamma\sqrt{T}]$, where $\alpha \geq 1/4$. As before, the total feedback of the best action $C_{i^\star,T}$ is distributed $Bin(T, \frac{1}{2} - \epsilon\delta)$. Thus, we have $\mathrm{E}[C_{i^\star,T}] = \frac{T}{2} - \epsilon\delta T$.

- if $\delta = \gamma\frac{1}{\epsilon}\sqrt{\frac{1}{T}}$ we have

  $$\mathrm{E}[C_{i^\star,T}] = \frac{T}{2} - \gamma\sqrt{T}$$

  Using the fact that for binomial distribution, $B(n, q)$, the median is $\lfloor nq \rfloor$ or $\lceil nq \rceil$ we have that with probability at least $\frac{1}{2}$

  $$C_{i^\star,T} \geq \frac{T}{2} - \gamma\sqrt{T}$$

- if $\delta = \frac{1}{2}$ we have that the distribution for the **ϵ-noisy feedback** of the *best action*, $c_{i^\star,t}$ is $B(\frac{1-\epsilon}{2})$, therefore

  $$\mathrm{E}[C_{i^\star,T}] = \frac{T}{2} - \frac{\epsilon}{2}T$$

  $\delta = \frac{1}{2}$ implies $\epsilon \leq 2\gamma\sqrt{\frac{1}{T}}$ (as $\delta = \min\{\gamma\frac{1}{\epsilon}\sqrt{\frac{1}{T}}, \frac{1}{2}\}$) thus,

  $$\frac{\epsilon}{2}T \leq T\gamma\sqrt{\frac{1}{T}} = \gamma\sqrt{T}$$

  Therefore, we still have that with probability at least $\frac{1}{2}$

  $$C_{i^\star,T} \geq \frac{T}{2} - \gamma\sqrt{T}$$

Putting all together with probability $\frac{\alpha}{2} \geq 1/8$ we have that $C_{i,T} < C_{i^\star,T}$.
From here the proof is similar to the proof for the case of $K \geq 27$ and yields,

$$\mathrm{E}[Regret(T)] \geq \frac{\alpha}{2}\delta T,$$

where $\delta = \min\{\delta = \gamma\frac{1}{\epsilon}\sqrt{\frac{1}{T}}, \frac{1}{2}\}$.

Proof of Theorem 6

By applying Lemma 3 and taking expectation on both sides we obtain

$$\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[\hat{\ell}_{i,t}] - \sum_{t=1}^{T}\mathrm{E}[\hat{\ell}_{k,t}] \leq \frac{\ln K}{\eta} + \eta\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[(\hat{\ell}_{i,t})^2] \tag{3}$$

Conditioning on $p_{i,t} \leq \frac{1-\theta}{2}$, the estimator $\hat{\ell}_{i,t}$ is biased, however, we can bound the deviation. Specifically,

$$\mathrm{E}[\hat{\ell}_{i,t}] = \theta * 0 + (1-\theta)\mathrm{E}[\hat{\ell}_{i,t} \mid p_{i,t} \leq \frac{1-\theta}{2}] = (1-\theta)\ell_{i,t}$$

This implies that

$$\ell_{i,t} - \theta \leq \mathrm{E}[\hat{\ell}_{i,t}] \leq \ell_{i,t}$$

To bound the second moment we have

$$\mathrm{E}[(\hat{\ell}_{i,t})^2] = \theta * 0 + (1-\theta)\mathrm{E}[(\hat{\ell}_{i,t})^2 \mid p_{i,t} \leq \frac{1-\theta}{2}] \leq \mathrm{E}[(\hat{\ell}_{i,t})^2 \mid p_{i,t} \leq \frac{1-\theta}{2}]$$

We bound the conditional expectation above as follows,

$$\begin{aligned}\mathrm{E}[(\hat{\ell}_{i,t})^2 \mid p_{i,t} \leq \frac{1-\theta}{2}] &= p_{i,t}\frac{(\bar{\ell}_{i,t} - p_{i,t})^2}{(1-2p_{i,t})^2} + (1-p_{i,t})\frac{(\ell_{i,t} - p_{i,t})^2}{(1-2p_{i,t})^2} \\ &\leq \frac{1}{(1-2p_{i,t})^2} = \frac{1}{\epsilon_{i,t}^2}\end{aligned}$$

Computing the expectation, given that the marginal is $U(0,1)$, we have,

$$\begin{aligned}\mathrm{E}[(\hat{\ell}_{i,t})^2] \leq \mathrm{E}[(\hat{\ell}_{i,t})^2 \mid p_{i,t} \leq \frac{1-\theta}{2}] &\leq \mathrm{E}_{\epsilon\sim U(0,1)}\,[\frac{1}{\epsilon^2}\mathbb{1}_{\epsilon\geq\theta}] \\ = \int_{\theta}^{1}\frac{1}{\epsilon^2}d\epsilon = -[\frac{1}{\epsilon}]_{\theta}^{1} &= \frac{1}{\theta} - 1 \leq \frac{1}{\theta}\end{aligned}$$

Bounding the expressions in inequality (3) we obtain

$$\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[\hat{\ell}_{i,t}] - \sum_{t=1}^{T}\mathrm{E}[\hat{\ell}_{k,t}] \geq \sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\ell_{i,t} - \sum_{t=1}^{T}\ell_{k,t} - \theta T$$

$$\frac{\ln K}{\eta} + \eta\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[(\hat{\ell}_{i,t})^2] \leq \frac{\ln K}{\eta} + \eta\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\frac{1}{\theta} = \frac{\ln K}{\eta} + \frac{\eta T}{\theta}$$

Rearranging the terms gives us,

$$\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\ell_{i,t} - \sum_{t=1}^{T}\ell_{k,t} \leq \frac{\ln K}{\eta} + \frac{\eta T}{\theta} + \theta T$$

Substituting $\eta = (\frac{\ln K}{T})^{2/3}$ and $\theta = (\frac{\ln K}{T})^{1/3}$ concludes the proof. ■

Proof of Theorem 7

Let $\theta = (\frac{\ln K}{T})^{1/3}$. Initially, the adversary choose an action $i^\star$ uniformly at random, and it will be the best action. Then, for each round t after observing ϵ_t, the adversary assigns losses as follow:

1. If $\epsilon_t \geq \theta$ then $\ell_{i,t} = 0$ for every action i.
2. Otherwise ($\epsilon_t < \theta$) the adversary draw a loss for each action as follows: for action $i^\star$ the loss is drawn from $B(\frac{1}{2} - \frac{1}{6})$ and for any other action $j \neq i^\star$ it is drawn from $B(\frac{1}{2})$.

Denote by T' the number of *bad rounds*. Since $\mathrm{E}[T'] = \theta T$ and the fact that for Binomial distribution, $B(n,p)$, the median is $\lfloor np \rfloor$ or $\lceil np \rceil$ we conclude that with probability at least $1/2$ we have $T' \geq \theta T$. Condition on this event we assume that $T' = \theta T$ (if $T' > \theta T$ we take the first θT rounds to be T') we reduce the *bad rounds* to the constant noise setting in the following way:
In the *bad rounds* we have $\epsilon_t \sim U(0,\theta)$. If we assume that in the *bad rounds* we have $\epsilon_t = \theta$, namely a constant noise, then we only reduced the noise in the model. We call the model with $\epsilon_t = \theta$ and $T = T'$ the *reduced model*. Therefore, a lower bound for the regret in the *reduced model* is also a lower bound for a model where $\epsilon_t \sim U(0,\theta)$.
Our *reduced model* is the **Full Information with Constant Noise** model with $T = T'$ and $\epsilon = \theta$. Denote by $Regret(T', \theta)$ the regret in the **Full Information with Constant Noise** model with horizon T' and noise parameter θ. Now, we can apply Theorem 5 on the *reduced model* and obtain that

$$Regret(T', \theta) \geq \frac{1}{24\theta}\sqrt{T' \ln K}$$

where $\gamma > 0$ is a constant. Setting $T' = \theta T = T^{2/3}(\ln K)^{1/3}$ we obtain that

$$Regret(\theta T, \theta) \geq \frac{1}{\theta}\sqrt{\theta T \ln K} = \frac{1}{24} T^{2/3}(\ln K)^{1/3}$$

Putting it back in the original model yields,

$$Regret(T) \geq \Pr[T' \geq \theta T] Regret(\theta T, \theta) \geq \frac{1}{2}\frac{1}{24} T^{2/3}(\ln K)^{1/3}$$

(We note that the number $\frac{1}{6}$ in the distribution $B(\frac{1}{2} - \frac{1}{6})$ the adversary uses, comes from the $\delta = \frac{1}{6\epsilon}\sqrt{\frac{\ln K}{T}}$ we use in the proof of Theorem 5 with $\epsilon = \theta$ and $T = T'$). ■

Proof of Theorem 8

We apply Lemma 3, and taking expectation on both sides, obtain,

$$\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[\hat{\ell}_{i,t}] - \sum_{t=1}^{T}\mathrm{E}[\hat{\ell}_{k,t}] \leq \frac{\ln K}{\eta} + \eta \sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[(\hat{\ell}_{i,t})^2] \tag{4}$$

Conditioning on $p_{i,t} \leq \frac{1-\theta}{2}$, the estimator $\hat{\ell}_{i,t}$ is biased, and we have

$$\mathrm{E}[\hat{\ell}_{i,t}] = F(\theta) * 0 + (1 - F(\theta))\mathrm{E}[\hat{\ell}_{i,t} \mid p_{i,t} \leq \frac{1-\theta}{2}] = (1 - F(\theta))\ell_{i,t}$$

This implies that

$$\ell_{i,t} - F(\theta) \leq \mathrm{E}[\hat{\ell}_{i,t}] \leq \ell_{i,t}$$

To bound the second moment we have

$$\mathrm{E}[(\hat{\ell}_{i,t})^2] = F(\theta) * 0 + (1 - F(\theta))\mathrm{E}[(\hat{\ell}_{i,t})^2 \mid p_{i,t} \leq \frac{1-\theta}{2}] \leq \mathrm{E}[(\hat{\ell}_{i,t})^2 \mid p_{i,t} \leq \frac{1-\theta}{2}]$$

We bound the above conditional expectation as follows,

$$\begin{aligned}\mathrm{E}[(\hat{\ell}_{i,t})^2 \mid p_{i,t} \leq \frac{1-\theta}{2}] &= p_{i,t}\frac{(\bar{\ell}_{i,t} - p_{i,t})^2}{(1-2p_{i,t})^2} + (1-p_{i,t})\frac{(\ell_{i,t} - p_{i,t})^2}{(1-2p_{i,t})^2} \\ &\leq \frac{1}{(1-2p_{i,t})^2} = \frac{1}{\epsilon_{i,t}^2}\end{aligned}$$

Bounding each side of inequality (4) we have

$$\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[\hat{\ell}_{i,t}] - \sum_{t=1}^{T}\mathrm{E}[\hat{\ell}_{k,t}] \geq \sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\ell_{i,t} - \sum_{t=1}^{T}\ell_{k,t} - F(\theta)T$$

$$\frac{\ln K}{\eta} + \eta\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[(\hat{\ell}_{i,t})^2] \leq \frac{\ln K}{\eta} + \eta\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[\frac{1}{\epsilon^2}\mathbb{1}_{\epsilon\geq\theta}]$$

Rearranging it all yield

$$\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\ell_{i,t} - \sum_{t=1}^{T}\ell_{k,t} \leq \frac{\ln K}{\eta} + \eta T g(\theta) + F(\theta)T$$

■

Proof of Corollary 9

Using Theorem 8 and the assumption we can write

$$Regret(T) \leq 2\sqrt{g(\theta)T\ln K} + \theta^{\alpha}T$$

since $g(\theta) = \mathrm{E}[\frac{1}{\epsilon^2}\mathbb{1}_{\epsilon\geq\theta}] \leq \frac{1}{\theta^2}$, we have,

$$Regret(T) \leq \frac{2}{\theta}\sqrt{T\ln K} + \theta^{\alpha}T$$

taking $\theta = (\frac{2}{\alpha})^{\frac{1}{1+\alpha}}(\frac{\ln K}{T})^{\frac{1}{2(1+\alpha)}}$ gives

$$Regret = O(T^{\frac{2+\alpha}{2+2\alpha}}(\ln K)^{\frac{\alpha}{2(1+\alpha)}})$$

■

Proof of Corollary 10

Applying Theorem 8 gives

$$Regret(T) \leq 2\sqrt{g(\theta)T \ln K} + F(\theta)T \tag{5}$$

To bound $g(\theta)$ we calculate

$$\begin{aligned} g(\theta) &= \mathrm{E}[\frac{1}{\epsilon^2}\mathbb{1}_{\epsilon \geq \theta}] \int_\theta^1 \frac{1}{\epsilon^2} \frac{\lambda}{1-e^{-\lambda}} e^{-\lambda\epsilon} d\epsilon \leq \frac{\lambda}{1-e^{-\lambda}} \int_\theta^1 \frac{1}{\epsilon^2} d\epsilon \\ &= \frac{\lambda}{1-e^{-\lambda}}(\frac{1}{\theta} - 1) \leq \frac{\lambda}{1-e^{-\lambda}} \frac{1}{\theta} \leq \frac{\lambda}{\theta} \end{aligned}$$

Bounding the second term we use the inequality $1 - e^{-x} \leq x$ for $x > 0$ and obtain

$$F(\theta) = \frac{\lambda}{1-e^{-\lambda}}(1 - e^{-\lambda\theta}) \leq \lambda^2\theta$$

Putting it back in (5) we have

$$Regret(T) \leq 2\sqrt{\frac{1}{\theta} \ln K} + \lambda^2\theta T$$

setting $\theta = \frac{1}{\lambda}(\frac{\ln K}{T})^{1/3}$ yields,

$$Regret(T) \leq 3\lambda T^{2/3} (\ln K)^{1/3}$$

■

Proof of Theorem 11

Let the number of actions be $K = 2$. Assume that initially the adversary picks the best action uniformly (that is, with probability $\frac{1}{2}$ action 1 will be the best action and with probability $\frac{1}{2}$ action 2 will be the best action). Let $i^\star \in \{1, 2\}$ be a random variable denoting the best action and $j = 3 - i^\star$ denote the worse action. On round t, after observing the noise parameters $p_{1,t}$ and $p_{2,t}$, the adversary selects the losses as follow:

1. For the best action, $i^\star$, the loss is drawn at every round independently from a Bernoulli r.v. with parameter 1/4, i.e., $\ell_{i^\star,t} \sim B(\frac{1}{4})$
2. For the worse action j: if $p_{j,t} < 1/4$ then the loss is $\ell_{j,t} = 0$, otherwise the loss is $\ell_{j,t} = 1$.

For the learner, observing the feedback $c_{i,t} = \ell_{i,t} \oplus r_{i,t}$, the loss of each action is a Bernoulli random variable. We will show that both actions will have the same probability of 1, namely 3/8, and therefore indistinguishable by the learner.

Now we calculate the expected value of the observed feedback, $c_{i,t} = \ell_{i,t} \oplus r_{i,t}$, for each action in a single round. We note that this expectation is taken over the draw of $\epsilon_{i,t} \sim U(0, 1)$, the draw $R_{i,t} \sim B(\frac{1-\epsilon_{i,t}}{2})$ and the draw of the losses $\ell_{i,t}$. We also note that if $\epsilon \sim U(0, 1)$ then $p \sim U(0, \frac{1}{2})$.

The expected loss of best action, $\ell_{i^\star,t}$ is drawn independently from the noise parameter $\epsilon_{i^\star,t}$ and the Bernoulli noise $R_{i,t}$. Therefore, we have

$$\begin{aligned}\mathrm{E}[c_{i^\star,t}] &= \mathrm{E}_p[\mathrm{E}_R[\mathrm{E}_\ell[\ell_{i^\star,t} \oplus R_{i^\star,t}] \mid p]] = \mathrm{E}_p[\mathrm{E}_R[\frac{1}{4}(1 \oplus R_{i^\star,t}) + \frac{3}{4}(0 \oplus R_{i^\star,t}) \mid p]] \\ &= \frac{1}{4}\mathrm{E}_p[p_{i^\star,t} \cdot 0 + (1 - p_{i^\star,t}) \cdot 1] + \frac{3}{4}\mathrm{E}_p[p_{i^\star,t} \cdot 1 + (1 - p_{i^\star,t}) \cdot 0] \\ &= \frac{1}{4} \cdot \frac{3}{4} + \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{8}\end{aligned}$$

For the worse action, action j, we have

$$\begin{aligned}\mathrm{E}[c_{j,t}] &= \mathrm{E}[\ell_{j,t} \oplus R_{j,t}] = \frac{1}{2}\mathrm{E}[0 \oplus R_{j,t} \mid p_{j,t} < 1/4] + \frac{1}{2}\mathrm{E}[1 \oplus R_{j,t} \mid \frac{1}{4} \le p_{j,t} < \frac{1}{2}] \\ &= \frac{1}{2}\mathrm{E}[p_{j,t} \mid p_{j,t} < \frac{1}{4}] + \frac{1}{2}\mathrm{E}[1 - p_{j,t} \mid \frac{1}{4} \le p_{j,t} < \frac{1}{2}] \\ &= \frac{1}{2} \cdot \frac{1}{8} + \frac{1}{2}(1 - \frac{3}{8}) = \frac{3}{8}\end{aligned}$$

This implies that the feedback of both the best and worse action is a Bernoulli random variable with parameter $\frac{3}{8}$, i.e., $B(\frac{3}{8})$. This clearly implies that the learner cannot distinguish between the two actions, and therefore, half the time it will select the worse action. The best action has an expected loss of $\frac{T}{4}$ while the worse action has a loss of $\frac{T}{2}$. This implies that the expected regret would be at least $\frac{T}{8}$. ■

Proof of Theorem 12

By applying Lemma 3 and taking expectation on both sides we obtain

$$\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[\hat{\ell}_{i,t}] - \sum_{t=1}^{T}\mathrm{E}[\hat{\ell}_{k,t}] \le \frac{\ln K}{\eta} + \eta\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[(\hat{\ell}_{i,t})^2]$$

Calculating the expectation of the estimator $\hat{\ell}_{i,t}$, and since $\ell_{i,t} \in \{0, 1\}$, we have,

$$\mathrm{E}[\hat{\ell}_{i,t}] = q_{i,t}\frac{1}{q_{i,t}}\mathrm{E}[c_{i,t}] = \mathrm{E}[c_{i,t}] = (1 - p)\ell_{i,t} + p\bar{\ell}_{i,t} = (1 - 2p)\ell_{i,t} + p = |\ell_{i,t} - p|$$

For the second moment, since $c_{i,t} \le 1$ we have

$$\mathrm{E}[(\ell_{i,t})^2] = q_{i,t}\frac{1}{q_{i,t}^2}\mathrm{E}[c_{i,t}] \le \frac{1}{q_{i,t}}$$

Putting things together we have

$$\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}|\ell_{i,t} - p| - \sum_{t=1}^{T}|\ell_{k,t} - p| \le \frac{\ln K}{\eta} + \eta TK \tag{6}$$

Using the notation of $\hat{L}_{ON,T} = \sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\hat{\ell}_{i,t}$, $\hat{L}_{k,T} = \sum_{t=1}^{T}\hat{\ell}_{k,t}$, $L_{ON,T} = \sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\ell_{i,t}$, and $L_{k,T} = \sum_{t=1}^{T}\ell_{k,t}$, we can write inequality (6) as

$$\mathrm{E}[\hat{L}_{ON,T}] - \mathrm{E}[\hat{L}_{k,T}] \le \frac{\ln K}{\eta} + \eta TK$$

Denote by $G_{t,b} = \{i \in A \mid \ell_{i,t} = b\}$ the set of actions with loss $b \in \{0,1\}$ in round t. Denote by $Q_t = \sum_{i \in G_{t,1}} q_{i,t}$ the distribution mass the learner gives actions in $G_{t,1}$. Using this notation we have $L_{ON,T} = \sum_{t=1}^{T} Q_t$. Now calculate the value of the estimated losses of the online algorithm,

$$\begin{aligned}
\mathrm{E}[\hat{L}_{ON,T}] &= \sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}|\ell_{i,t} - p| = \sum_{t=1}^{T}[p \sum_{i \in G_{t,0}} q_{i,t} + (1-p) \sum_{i \in G_{t,1}} q_{i,t}] \\
&= \sum_{t=1}^{T}[p(1-Q_t) + (1-p)Q_t] = \sum_{t=1}^{T}[p + (1-2p)Q_t] \\
&= (1-2p)L_{ON,T} + pT
\end{aligned}$$

Similarly, for the term $\mathrm{E}[\hat{L}_{k,T}]$ we have,

$$\begin{aligned}
\mathrm{E}[\hat{L}_{k,T}] &= \sum_{t=1}^{T} |\ell_{k,t} - p| = \sum_{t|\ell_{t,k}=0} p + \sum_{t|\ell_{t,k}=1} (1-p) \\
&= p(T - L_{k,T}) + (1-p)L_{k,T} = (1-2p)L_{k,T} + pT
\end{aligned}$$

Putting all together,

$$\mathrm{E}[\hat{L}_{ON,T}] - \mathrm{E}[\hat{L}_{k,T}] = (1-2p)L_{ON,T} + pT - [(1-2p)L_{k,T} + pT] = (1-2p)[L_{ON,T} - L_{k,T}]$$

Dividing by both sides of inequity by $(1-2p)$ and using $\eta = \sqrt{\frac{\ln K}{TK}}$ we obtain that

$$Regret(T) = L_{ON,T} - \min_{k \in A} L_{k,T} \leq \frac{1}{1-2p}(\frac{\ln K}{\eta} + \eta TK) = \frac{2}{\epsilon}\sqrt{TK \ln K}$$

■

Proof of Theorem 13

We first define K different problem instances, one per action. Let $\beta \in (0,1)$ be a parameter. We denote by J_i the problem instance where action i loss is drawn from the distribution $B(\frac{1-\beta}{2})$ while the other actions loss is drawn from the distribution $B(\frac{1}{2})$. For problem instance J_i, we refer action i as the *best action*. The proof will show that in some sense those instances are indistinguishable for any algorithm.

For the proof, we will think of the online algorithm as a leaner making "prediction" for the best action at each round t. The main part of the proof is to show that if T is not large enough the algorithm has to have a constant mistake rate.

We denote by $\Pr[I_t = i|J_i]$ the probability that in instance J_i, at round t the algorithm selects action i (the best action in instance J_i). The following lemma shows that for many actions the algorithm will make a mistake.

Lemma 6 *Consider a deterministic algorithm for the Bandit with Constant Noise problem with noise $p = \frac{1-\epsilon}{2}$. There exist a constant γ such that if $t < \gamma \frac{K}{\epsilon^2\beta^2}$ then there exist at least $\lceil \frac{K}{2} \rceil$ actions i such that*

$$\Pr[I_t = i|J_i] < \frac{3}{4}$$

Proof Consider the feedback distribution for each problem instance J_j and action i. First, if $\ell_{i,t} \sim B(\frac{1}{2})$ then $c_{i,t} \sim B(\frac{1}{2})$ (the noise does not have any influence). For the *best action*, i.e., j, we have $c_{j,t} \sim B(\frac{1-\epsilon\beta}{2})$ since

$$\begin{aligned} Pr[c_{j,t}=1] &= Pr[\ell_{j,t}=1]Pr[R_\epsilon=0] + Pr[\ell_{j,t}=0]Pr[R_\epsilon=1] \\ &= (\frac{1-\beta}{2})(\frac{1+\epsilon}{2}) + (\frac{1+\beta}{2})(\frac{1-\epsilon}{2}) = \frac{1-\epsilon\beta}{2} \end{aligned}$$

Applying *Lemma 2.10* of Slivkins (2017) on the feedbacks $c_{i,t}$ completes the proof. ■

Corollary 7 *Choose the best action $i^\star$ uniformly from A and use instance $J_{i^\star}$. For any algorithm, for any round $t < \gamma\frac{K}{\epsilon^2\beta^2}$, we have* $\Pr[I_t \neq i^\star] \geq 1/8$.

Proof For a deterministic algorithm the corollary follows since by Lemma 6 with probability at least $\frac{1}{2}$ the selected $i^\star$ is such that $\Pr[I_t \neq i^\star | J_{i^\star}] \geq \frac{1}{4}$. Since a randomized algorithm is a distribution over deterministic algorithms that claim hold also for randomized algorithms. ■

Proof of Theorem 13: Let $\beta = \min\{\frac{\sqrt{\gamma}}{\epsilon}\sqrt{\frac{K}{T}}, 1\}$. By Corollary 7, we have that in each round t

$$\Pr[I_t \neq i^\star] \geq \frac{1}{8}$$

Denote by $\Delta_t = \mathrm{E}[\ell_{I_t,t}] - \mathrm{E}[\ell_{i^\star,t}]$ the regret of round t. Note that if $I_t \neq i^\star$ then $\Delta_t = \frac{1}{2} - \frac{1-\beta}{2} = \frac{\beta}{2}$. Therefore, the expected regret at round t is

$$\mathrm{E}[\Delta_t] = \Pr[I_t \neq i^\star]\frac{\beta}{2}$$

Summing over the rounds we have,

$$Regret(T) = \sum_{t=1}^{T} \mathrm{E}[\Delta_t] \geq \frac{1}{16}\beta T$$

Since $\beta = \min\{\frac{\sqrt{\gamma}}{\epsilon}\sqrt{\frac{K}{T}}, 1\}$, we have

$$Regret(T) \geq \min\{\frac{\sqrt{\gamma}}{16}\frac{1}{\epsilon}\sqrt{TK}, \frac{1}{16}T\}$$

■

Proof of Theorem 14

By applying Lemma 3 and taking expectation on both sides we obtain

$$\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[\hat{\ell}_{i,t}] - \sum_{t=1}^{T}\mathrm{E}[\hat{\ell}_{k,t}] \leq \frac{\ln K}{\eta} + \eta\sum_{t=1}^{T}\sum_{i=1}^{K} q_{i,t}\mathrm{E}[(\hat{\ell}_{i,t})^2] \tag{7}$$

Conditioning on $p_{i,t} \leq \frac{1-\theta}{2}$, the estimator $\hat{\ell}_{i,t}$ is unbiased, since

$$\mathrm{E}[\hat{\ell}_{i,t} \;\mid\; p_t \leq \frac{1-\theta}{2}] = q_{i,t}[\frac{1}{q_{i,t}} \frac{p\bar{\ell}_{i,t} + (1-p)\ell_{i,t} - p}{1-2p}] = \ell_{i,t} \; .$$

However, overall the estimator is biased,

$$\mathrm{E}[\hat{\ell}_{i,t}] = \theta * 0 + (1-\theta)\mathrm{E}[\hat{\ell}_{i,t} \mid p_{i,t} \leq \frac{1-\theta}{2}] = (1-\theta)\ell_{i,t}$$

This implies that

$$\ell_{i,t} - \theta \leq \mathrm{E}[\hat{\ell}_{i,t}] \leq \ell_{i,t}$$

To bound the second moment we have

$$\mathrm{E}[(\hat{\ell}_{i,t})^2] = \theta * 0 + (1-\theta)\mathrm{E}[(\hat{\ell}_{i,t})^2 \mid p_{i,t} \leq \frac{1-\theta}{2}] \leq \mathrm{E}[(\hat{\ell}_{i,t})^2 \mid p_{i,t} \leq \frac{1-\theta}{2}]$$

The conditional expectation of the second moment is bounded as follows,

$$\mathrm{E}[(\hat{\ell}_{i,t})^2 \mid p_t \leq \frac{1-\delta}{2}] = \frac{1}{q_{i,t}}[p_t \frac{(\bar{\ell}_{i,t} - p_t)^2}{(1-2p_t)^2} + (1-p_t)\frac{(\ell_{i,t} - p_t)^2}{(1-2p_t)^2}] \leq \frac{1}{q_{i,t}} \frac{1}{(1-2p_t)^2} = \frac{1}{q_{i,t}} \frac{1}{\epsilon_t^2}$$

Since the marginal of the noise distribution D is uniform, we have,

$$\begin{aligned} \mathrm{E}[(\hat{\ell}_{i,t})^2] &\leq \mathrm{E}[(\hat{\ell}_{i,t})^2 \mid p_{i,t} \leq \frac{1-\theta}{2}] \leq \mathrm{E}_{\epsilon \sim U(0,1)}\, [\frac{1}{q_{i,t}} \frac{1}{\epsilon^2} \mathbb{1}_{\epsilon \geq \theta}] \\ &= \frac{1}{q_{i,t}} \int_\theta^1 \frac{1}{\epsilon^2} d\epsilon = -\frac{1}{q_{i,t}} [\frac{1}{\epsilon}]_\theta^1 = \frac{1}{q_{i,t}} (\frac{1}{\theta} - 1) \leq \frac{1}{q_{i,t}} \frac{1}{\theta} \end{aligned} \tag{8}$$

Bounding each side of inequality (7) we have

$$\begin{aligned} &\sum_{t=1}^{T} \sum_{i=1}^{K} q_{i,t} \mathrm{E}[\hat{\ell}_{i,t}] - \sum_{t=1}^{T} \mathrm{E}[\hat{\ell}_{k,t}] \geq \sum_{t=1}^{T} \sum_{i=1}^{K} q_{i,t}\ell_{i,t} - \sum_{t=1}^{T} \ell_{k,t} - \theta T \\ &\frac{\ln K}{\eta} + \eta \sum_{t=1}^{T} \sum_{i=1}^{K} q_{i,t} \mathrm{E}[(\hat{\ell}_{i,t})^2] \leq \frac{\ln K}{\eta} + \eta \sum_{t=1}^{T} \sum_{i=1}^{K} q_{i,t} [\frac{1}{q_{i,t}} \frac{1}{\theta}] = \frac{\ln K}{\eta} + \frac{\eta T K}{\theta} \end{aligned} \tag{9}$$

Rearranging it all yield

$$\sum_{t=1}^{T} \sum_{i=1}^{K} q_{i,t}\ell_{i,t} - \sum_{t=1}^{T} \ell_{k,t} \leq \frac{\ln K}{\eta} + \frac{\eta T K}{\theta} + \theta T$$

Substituting $\eta = \frac{(\ln K)^{2/3}}{K^{1/3} T^{2/3}}$ and $\theta = \frac{K^{1/3} (\ln K)^{1/3}}{T^{1/3}}$ concludes the proof. ■

Proof of Theorem 15

Let $\theta = (\frac{K}{T})^{1/3}$. Initially, the adversary choose an action $i^\star$ uniformly at random, and it will be the best action. Then, for each round t after observing ϵ_t, the adversary assigns losses as follow: fix $\beta = \frac{\sqrt{\gamma}}{\theta}\sqrt{\frac{K}{T}} = \sqrt{\gamma}(\frac{K}{T})^{1/6}$ and at round t do

1. if $\epsilon_t \geq \theta$ then $\ell_{i,t} = 0$ for every action i.
2. Otherwise ($\epsilon_t < \theta$) the adversary draw a loss for each action as follows: for action $i^\star$ the loss is drawn from $B(\frac{1}{2} - \beta)$ and for any other action $j \neq i^\star$ it is drawn from $B(\frac{1}{2})$.

Denote by T' the number of *bad rounds*. Since $\mathrm{E}[T'] = \theta T$ and the fact that for Binomial distribution, $B(n,p)$, the median is $\lfloor np \rfloor$ or $\lceil np \rceil$ we conclude that with probability at least $1/2$ we have $T' \geq \theta T$. Condition on this event we assume that $T' = \theta T$ (if $T' > \theta T$ we take the first θT rounds to be T') we reduce the *bad rounds* to the constant noise setting in the following way:
In the *bad rounds* we have $\epsilon_t \sim U(0,\theta)$. If we assume that in the *bad rounds* we have $\epsilon_t = \theta$, namely a constant noise, then we only reduced the noise in the model. We call the model with $\epsilon_t = \theta$ and $T = T'$ the *reduced model*. Therefore, a lower bound for the regret in the *reduced model* is also a lower bound for a model where $\epsilon_t \sim U(0,\theta)$.
Our *reduced model* is the **Bandit with Constant Noise** model with $T = T'$ and $\epsilon = \theta$. Denote by $Regret(T', \theta)$ the regret in the **Bandit with Constant Noise** model with horizon T' and noise parameter θ. Now, we can apply Theorem 13 on the *reduced model* and obtain that

$$Regret(T', \theta) \geq \gamma \frac{1}{\theta}\sqrt{T'K}$$

where $\gamma > 0$ is a constant. Setting $T' = \theta T = T^{2/3}K^{1/3}$ we obtain that

$$Regret(\theta T, \theta) \geq \frac{1}{\theta}\sqrt{\theta T K} = \gamma T^{2/3}K^{1/3}$$

Putting it back in the original model yields,

$$Regret(T) \geq \Pr[T' \geq \theta T] * Regret(\theta T, \theta) \geq \frac{\gamma}{2}T^{2/3}K^{1/3}$$

(We note here that the choice of β is according to the proof of Theorem 13 with $\epsilon = \theta$ and $T = T' = \theta T$). ■

References

Philip Klein and Neal Young. On the number of iterations for dantzig-wolfe optimization and packing-covering approximation algorithms. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 320–327. Springer, 1999.

Aleksandrs Slivkins. Introduction to multi-armed bandits, 2017. URL `http://slivkins.com/work/MAB-book.pdf`.