Explore no more

Improved high-probability regret bounds for non-stochastic bandits

Gergely Neu

INRIA Lille, SequeL team

→ Universitat Pompeu Fabra, Barcelona

Outline

- Non-stochastic bandits
- The EXP3 algorithm
- High-probability bounds
- **Simple** high-probability bounds

Non-stochastic bandits

For all rounds $t = 1,2,\ldots,T$

- Learner chooses action/arm $I_t \in \{1,2,\ldots,K\}$
- Environment chooses loss function $\ell_{t,i} \in [0,1]^K$
- Learner suffers and observes loss ℓ_{t,I_t}

Non-stochastic bandits

For all rounds $t = 1,2,\ldots,T$

- Learner chooses action/arm $I_t \in \{1,2,\ldots,K\}$
- Environment chooses loss function $\ell_{t,i} \in [0,1]^K$
- Learner suffers and observes loss ℓ_{t,I_t}

Goal: minimize regret

$$R_T = \sum_{t=1}^{T} \ell_{t,I_t} - \min_{i \in \{1,2,\ldots,N\}} \sum_{t=1}^{T} \ell_{t,i}$$

Various regret notions

- Pseudo-regret or weak regret:

$$\hat{R}_T = \mathbf{E}\left[\sum_{t=1}^{T} \ell_{t,I_t}\right] - \min_{i\in\{1,2,\ldots,N\}} \mathbf{E}\left[\sum_{t=1}^{T} \ell_{t,i}\right]$$

- Expected regret:

$$\mathbf{E}[R_T] = \mathbf{E}\left[\sum_{t=1}^{T} \ell_{t,I_t}\right] - \mathbf{E}\left[\min_{i\in\{1,2,\ldots,N\}} \sum_{t=1}^{T} \ell_{t,i}\right]$$

- Regret:

$$R_T = \sum_{t=1}^{T} \ell_{t,I_t} - \min_{i\in\{1,2,\ldots,N\}} \sum_{t=1}^{T} \ell_{t,i}$$

Various regret notions

- Pseudo-regret or weak regret:

$$\hat{R}_T = \mathbf{E}\left[\sum_{t=1}^{T} \ell_{t,I_t}\right] - \min_{i\in\{1,2,\ldots,N\}} \mathbf{E}\left[\sum_{t=1}^{T} \ell_{t,i}\right]$$

- Expected regret:

$$\mathbf{E}[R_T] = \mathbf{E}\left[\sum_{t=1}^{T} \ell_{t,I_t}\right] - \mathbf{E}\left[\min_{i\in\{1,2,\ldots,N\}} \sum_{t=1}^{T} \ell_{t,i}\right]$$

- Regret:

$$R_T = \sum_{t=1}^{T} \ell_{t,I_t} - \min_{i\in\{1,2,\ldots,N\}} \sum_{t=1}^{T} \ell_{t,i}$$

Various regret notions

- Pseudo-regret or weak regret:

$$\hat{R}_T = \mathbf{E}\left[\sum_{t=1}^{T} \ell_{t,I_t}\right] - \min_{i\in\{1,2,\ldots,N\}} \mathbf{E}\left[\sum_{t=1}^{T} \ell_{t,i}\right]$$

- Expected regret:

$$\mathbf{E}[R_T] = \mathbf{E}\left[\sum_{t=1}^{T} \ell_{t,I_t}\right] - \mathbf{E}\left[\min_{i\in\{1,2,\ldots,N\}} \sum_{t=1}^{T} \ell_{t,i}\right]$$

- Regret:

$$R_T = \sum_{t=1}^{T} \ell_{t,I_t} - \min_{i\in\{1,2,\ldots,N\}} \sum_{t=1}^{T} \ell_{t,i}$$

A classic algorithm

EXP3 (Auer, Cesa-Bianchi, Freund and Schapire, 1995, 2002)

Parameters: $\eta > 0$, $\gamma \in [0,1]$

Initialization: For all i, set $w_{1,i} = 0$

For all rounds $t = 1,2,\ldots,T$

- For all i, let

$$p_{t,i} = (1-\gamma)\frac{w_{t,i}}{\sum_{j=1}^{K} w_{t,j}} + \frac{\gamma}{K}$$

- Sample $I_t \sim \boldsymbol{p}_t$
- For all i, let

$$\hat{r}_{t,i} = \frac{r_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}}$$

- For all i, update weights as

$$w_{t+1,i} = w_{t,i} \cdot e^{\eta \hat{r}_{t,i}}$$

A classic algorithm

EXP3 (Auer, Cesa-Bianchi, Freund and Schapire, 1995, 2002)

Parameters: $\eta > 0$, $\gamma \in [0,1]$

Initialization: For all i, set $w_{1,i} = 0$

For all rounds $t = 1,2,\dots,T$

- For all i, let

$$p_{t,i} = (1-\gamma)\frac{w_{t,i}}{\sum_{j=1}^{K} w_{t,j}} + \frac{\gamma}{K}$$

- Sample $I_t \sim \boldsymbol{p}_t$
- For all i, let

$$\hat{r}_{t,i} = \frac{r_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}}$$

- For all i, update weights as

$$w_{t+1,i} = w_{t,i} \cdot e^{\eta \hat{r}_{t,i}}$$

"reward" or "gain":
$r_{t,i} = 1 - \ell_{t,i}$

A classic algorithm

EXP3 (Auer, Cesa-Bianchi, Freund and Schapire, 1995, 2002)

Parameters: $\eta > 0, \gamma \in [0,1]$

Initialization: For all i, set $w_{1,i} = 0$

For all rounds $t = 1,2,\ldots,T$

- For all i, let
$$p_{t,i} = (1-\gamma)\frac{w_{t,i}}{\sum_{j=1}^{K} w_{t,j}} + \frac{\gamma}{K}$$
- Sample $I_t \sim \boldsymbol{p}_t$
- For all i, let
$$\hat{r}_{t,i} = \frac{r_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}}$$
- For all i, update weights as
$$w_{t+1,i} = w_{t,i} \cdot e^{\eta \hat{r}_{t,i}}$$

"reward" or "gain": $r_{t,i} = 1 - \ell_{t,i}$

Theorem: When tuned properly, EXP3 guarantees $\hat{R}_T \leq 2.63\sqrt{KT \log K}$

An improved version

EXP3 (e.g., Stoltz, 2005, Bubeck and Cesa-Bianchi, 2012)

Parameters: $\eta > 0, \gamma \in [0,1]$

Initialization: For all i, set $w_{1,i} = 0$

For all rounds $t = 1,2,\ldots,T$

- For all i, let

$$p_{t,i} = (1-\gamma)\frac{w_{t,i}}{\sum_{j=1}^{K} w_{t,j}} + \frac{\gamma}{K}$$

- Sample $I_t \sim \boldsymbol{p}_t$
- For all i, let

$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}}$$

- For all i, update weights as

$$w_{t+1,i} = w_{t,i} \cdot e^{-\eta \hat{\ell}_{t,i}}$$

An improved version

EXP3 (e.g., Stoltz, 2005, Bubeck and Cesa-Bianchi, 2012)

Parameters: $\eta > 0$, $\gamma \in [0,1]$

Initialization: For all i, set $w_{1,i} = 0$

For all rounds $t = 1,2, \dots, T$

- For all i, let

$$p_{t,i} = (1-\gamma)\frac{w_{t,i}}{\sum_{j=1}^{K} w_{t,j}} + \frac{\gamma}{K}$$

- Sample $I_t \sim \boldsymbol{p}_t$
- For all i, let

$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}}$$

- For all i, update weights as

$$w_{t+1,i} = w_{t,i} \cdot e^{-\eta \hat{\ell}_{t,i}}$$

Theorem: When tuned properly, EXP3 guarantees $\hat{R}_T \le \sqrt{2KT \log K}$

An improved version

EXP3 (e.g., Stoltz, 2005, Bubeck and Cesa-Bianchi, 2012)

Parameters: $\eta > 0, \gamma \in [0,1]$

Initialization: For all i, set $w_{1,i} = 0$

For all rounds $t = 1,2,\ldots,T$

- For all i, let

$$p_{t,i} = (1-\gamma)\frac{w_{t,i}}{\sum_{j=1}^{K} w_{t,j}} + \frac{\gamma}{K}$$

- Sample $I_t \sim \boldsymbol{p}_t$
- For all i, let

$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}}$$

- For all i, update weights as

$$w_{t+1,i} = w_{t,i} \cdot e^{-\eta \hat{\ell}_{t,i}}$$

The best choice is $\gamma = 0$

Theorem: When tuned properly, EXP3 guarantees $\hat{R}_T \leq \sqrt{2KT \log K}$

An improved version

EXP3 (e.g., Stoltz, 2005, Bubeck and Cesa-Bianchi, 2012)

Parameters: $\eta > 0, \gamma \in [0,1]$

Initialization: For all i, set $w_{1,i} = 0$

For all rounds $t = 1,2, \dots, T$

- For all i, let

$$p_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^{K} w_{t,j}}$$

- Sample $I_t \sim \boldsymbol{p}_t$
- For all i, let

$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}}$$

- For all i, update weights as

$$w_{t+1,i} = w_{t,i} \cdot e^{-\eta \hat{\ell}_{t,i}}$$

The best choice is $\gamma = 0$

Theorem: When tuned properly, EXP3 guarantees $\hat{R}_T \leq \sqrt{2KT \log K}$

Why explore then?

Conjecture (folklore):
Setting $\gamma > 0$ is necessary to prove strong bounds on R_T that hold **with high probability**

Why explore then?

Conjecture (folklore):
Setting $\gamma > 0$ is necessary to prove strong bounds on R_T that hold **with high probability**

FALSE

A high-confidence algorithm

EXP3 (Auer, Cesa-Bianchi, Freund and Schapire, 1995, 2002)

Parameters: $\eta > 0, \gamma \in [0,1]$

Initialization: For all i, set $w_{1,i} = 0$

For all rounds $t = 1,2, \ldots, T$

- For all i, let
$$p_{t,i} = (1-\gamma)\frac{w_{t,i}}{\sum_{j=1}^{K} w_{t,j}} + \frac{\gamma}{K}$$
- Sample $I_t \sim \boldsymbol{p}_t$
- For all i, let
$$\hat{r}_{t,i} = \frac{r_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}}$$
- For all i, update weights as
$$w_{t+1,i} = w_{t,i} \cdot e^{\eta \hat{r}_{t,i}}$$

A high-confidence algorithm

EXP3.P (Auer, Cesa-Bianchi, Freund and Schapire, 1995, 2002)

Parameters: $\eta > 0$, $\gamma \in [0,1]$, $\beta > 0$

Initialization: For all i, set $w_{1,i} = 0$

For all rounds $t = 1,2, \dots, T$

- For all i, let

$$p_{t,i} = (1-\gamma)\frac{w_{t,i}}{\sum_{j=1}^{K} w_{t,j}} + \frac{\gamma}{K}$$

- Sample $I_t \sim \boldsymbol{p}_t$
- For all i, let

$$\hat{r}_{t,i} = \frac{r_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}} + \frac{\beta}{p_{t,i}}$$

- For all i, update weights as

$$w_{t+1,i} = w_{t,i} \cdot e^{\eta \hat{r}_{t,i}}$$

A high-confidence algorithm

EXP3.P (Auer, Cesa-Bianchi, Freund and Schapire, 1995, 2002)

Parameters: $\eta > 0$, $\gamma \in [0,1]$, $\beta > 0$

Initialization: For all i, set $w_{1,i} = 0$

For all rounds $t = 1,2,\ldots,T$

- For all i, let

$$p_{t,i} = (1-\gamma)\frac{w_{t,i}}{\sum_{j=1}^{K} w_{t,j}} + \frac{\gamma}{K}$$

- Sample $I_t \sim \boldsymbol{p}_t$
- For all i, let

$$\hat{r}_{t,i} = \frac{r_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}} + \frac{\beta}{p_{t,i}}$$

- For all i, update weights as

$$w_{t+1,i} = w_{t,i} \cdot e^{\eta \hat{r}_{t,i}}$$

Theorem: When tuned properly, EXP3.P guarantees

$$R_T \leq 5.25\sqrt{KT\log(K/\delta)}$$

w.p. $\geq 1-\delta$

A high-confidence algorithm

EXP3.P (Auer, Cesa-Bianchi, Freund and Schapire, 1995, 2002)

Parameters: $\eta > 0$, $\gamma \in [0,1]$, $\beta > 0$

Initialization: For all i, set $w_{1,i} = 0$

For all rounds $t = 1,2, \dots, T$

- For all i, let

$$p_{t,i} = (1-\gamma)\frac{w_{t,i}}{\sum_{j=1}^{K} w_{t,j}} + \frac{\gamma}{K}$$

- Sample $I_t \sim \boldsymbol{p}_t$
- For all i, let

$$\hat{r}_{t,i} = \frac{r_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}} + \frac{\beta}{p_{t,i}}$$

- For all i, update weights as

$$w_{t+1,i} = w_{t,i} \cdot e^{\eta \hat{r}_{t,i}}$$

This requires setting $\gamma = \Theta\left(\sqrt{K/T}\right)$

Theorem: When tuned properly, EXP3.P guarantees $R_T \leq 5.25\sqrt{KT \log(K/\delta)}$ w.p. $\geq 1-\delta$

Why not use losses then?

- For some reason…
  - Cesa-Bianchi and Lugosi (2006)

Two modifications of the strategy described at the beginning of the previous section are necessary to achieve better rates of convergence. First of all, the modified strategy estimates *gains* instead of losses. For convenience, we introduce the notation $g(i, Y_t) = 1 - \ell(i, Y_t)$ and the estimated gains

  - Bubeck and Cesa-Bianchi (2012)

This issue can be solved by combining the mixing idea with a different estimate for losses. In fact, the core idea is more transparent when expressed in terms of gains, and so we turn to the gain version of the problem. The trick is to introduce a bias in the gain estimate which allows to derive a high probability statement on this estimate.

  - + ALL lecture notes

The main challenge

$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}}$$

- Unbiased ☺
- Has HUGE variance ☹

The main challenge

$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}}$$

- Unbiased ☺
- Has HUGE variance ☹

$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}} - \frac{\beta}{p_{t,i}}$$

- Biased 😐
- Variance under control ☺
- … only works if $\gamma > 0$ ☹

The main challenge

$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}}$$

- Unbiased ☺
- Has HUGE variance ☹

$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}_{\{I_t = i\}} - \frac{\beta}{p_{t,i}}$$

- Biased 😐
- Variance under control ☺
- … only works if $\gamma > 0$ ☹

This work: Implicit eXploration (IX)

$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i} + \gamma} \mathbf{1}_{\{I_t = i\}}$$

- Biased 😐
- Variance under control ☺
- “Pretend to explore”

See also Kocák, Neu, Valko and Munos (NIPS 2014), Neu (COLT 2015)

The algorithm

EXP3-IX

Parameters: $\eta > 0$, $\gamma > 0$

Initialization: For all i, set $w_{1,i} = 0$

For all rounds $t = 1,2, \dots, T$

- For all i, let
$$p_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^{K} w_{t,j}}$$
- Sample $I_t \sim \boldsymbol{p}_t$
- For all i, let
$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i} + \gamma} \mathbf{1}_{\{I_t = i\}}$$
- For all i, update weights as
$$w_{t+1,i} = w_{t,i} \cdot e^{-\eta \hat{\ell}_{t,i}}$$

The algorithm

EXP3-IX

Parameters: $\eta > 0$, $\gamma > 0$

Initialization: For all i, set $w_{1,i} = 0$

For all rounds $t = 1,2, \dots, T$

- For all i, let
$$p_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^{K} w_{t,j}}$$
- Sample $I_t \sim \boldsymbol{p}_t$
- For all i, let
$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i} + \gamma} \mathbf{1}_{\{I_t = i\}}$$
- For all i, update weights as
$$w_{t+1,i} = w_{t,i} \cdot e^{-\eta \hat{\ell}_{t,i}}$$

No explicit exploration!

The algorithm

EXP3-IX

Parameters: $\eta > 0$, $\gamma > 0$

Initialization: For all i, set $w_{1,i} = 0$

For all rounds $t = 1,2, \dots, T$

- For all i, let
$$p_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^{K} w_{t,j}}$$
- Sample $I_t \sim \boldsymbol{p}_t$
- For all i, let
$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i} + \gamma} \mathbf{1}_{\{I_t = i\}}$$
- For all i, update weights as
$$w_{t+1,i} = w_{t,i} \cdot e^{-\eta \hat{\ell}_{t,i}}$$

No explicit exploration!

Theorem: When tuned properly, EXP3-IX guarantees
$$R_T \leq 2\sqrt{2KT \log(K/\delta)}$$
w.p. $\geq 1 - \delta$

The main tool

Lemma: With probability at least $1 - \delta$,

$$\sum_{t=1}^{T} (\hat{\ell}_{t,i} - \ell_{t,i}) \leq \frac{\log(K/\delta)}{2\gamma}$$

holds simultaneously for all $i \in \{1,2, \dots K\}$.

The main tool

Lemma: With probability at least $1-\delta$,
$$\sum_{t=1}^{T}(\hat{\ell}_{t,i}-\ell_{t,i}) \leq \frac{\log(K/\delta)}{2\gamma}$$
holds simultaneously for all $i \in \{1,2,\dots K\}$.

Proof idea:
$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}+\gamma}\mathbf{1}_{\{I_t=i\}} = \frac{\ell_{t,i}}{p_{t,i}+\gamma}\left(\mathbf{1}_{\{I_t=i\}}+\gamma\right) - \frac{\gamma\ell_{t,i}}{p_{t,i}+\gamma}$$

The main tool

Lemma: With probability at least $1-\delta$,

$$\sum_{t=1}^{T}\left(\widehat{\ell}_{t,i}-\ell_{t,i}\right) \leq \frac{\log(K/\delta)}{2\gamma}$$

holds simultaneously for all $i \in \{1,2,\dots K\}$.

Proof idea:

$$\widehat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}+\gamma}\mathbf{1}_{\{I_t=i\}} = \frac{\ell_{t,i}}{p_{t,i}+\gamma}\left(\mathbf{1}_{\{I_t=i\}}+\gamma\right) - \frac{\gamma\ell_{t,i}}{p_{t,i}+\gamma}$$

$$\approx \frac{\ell_{t,i}}{p_{t,i}}\mathbf{1}_{\{I_t=i\}} - \frac{\beta}{p_{t,i}}$$

And then use Freedman's inequality as done by, e.g., Beygelzimer et al. (2011)

A more elementary proof

Appendix A. The proof of Lemma 1

Fix any t. For convenience, we will use the notation $\beta = 2\gamma$. First, observe that for any i,

$$\widetilde{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i} + \gamma} \mathbb{1}_{\{I_t = i\}} \leq \frac{\ell_{t,i}}{p_{t,i} + \gamma \ell_{t,i}} \mathbb{1}_{\{I_t = i\}} = \frac{1}{2\gamma} \cdot \frac{2\gamma \ell_{t,i}/p_{t,i}}{1 + \gamma \ell_{t,i}/p_{t,i}} \mathbb{1}_{\{I_t = i\}} \leq \frac{1}{\beta} \cdot \log\left(1 + \beta \widehat{\ell}_{t,i}\right),$$

where the first step follows from $\ell_{t,i} \leq 1$ and last one from the elementary inequality $\frac{z}{1+z/2} \leq \log(1+z)$ that holds for all $z \geq 0$.

Let us define the notations $\widetilde{\lambda}_t = \sum_{i=1}^K h_{t,i} \widetilde{\ell}_{t,i}$ and $\lambda_t = \sum_{i=1}^K h_{t,i} \ell_{t,i}$. Then, we get that

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\beta \widetilde{\lambda}_t\right) \middle| \mathcal{F}_{t-1}\right] &\leq \mathbb{E}\left[\exp\left(\sum_{i=1}^K h_{t,i} \cdot \log\left(1 + \beta \widehat{\ell}_{t,i}\right)\right) \middle| \mathcal{F}_{t-1}\right] \\
&\leq \mathbb{E}\left[\prod_{i=1}^K \left(1 + \beta \widehat{\ell}_{t,i}\right)^{h_{t,i}} \middle| \mathcal{F}_{t-1}\right] \leq \prod_{i=1}^K \left(1 + \beta \ell_{t,i}\right)^{h_{t,i}} \\
&\leq \exp\left(\beta \sum_{i=1}^K h_{t,i} \ell_{t,i}\right) = \exp\left(\beta \lambda_t\right),
\end{aligned}$$

where we used Jensen's inequality for the concave function $z^{h_{t,i}}$ and $\mathbb{E}\left[\widehat{\ell}_{t,i} \middle| \mathcal{F}_{t-1}\right] \leq \ell_{t,i}$ that holds by definition of the loss estimates, and $1 + z \leq e^z$ that holds for all $z \in \mathbb{R}$. As a result, the process $W_t = \exp\left(\beta \sum_{s=1}^t (\widetilde{\lambda}_s - \lambda_s)\right)$ is a supermartingale with respect to $(\mathcal{F}_t)$: $\mathbb{E}\left[W_t \middle| \mathcal{F}_{t-1}\right] \leq W_{t-1}$. Observe that, since $W_0 = 1$, this implies $\mathbb{E}\left[W_t\right] \leq \mathbb{E}\left[W_{t-1}\right] \leq \ldots \leq 1$, and thus by Markov's inequality,

$$\mathbb{P}\left[\sum_{s=1} (\widetilde{\lambda}_s - \lambda_s) > \varepsilon\right] \leq \mathbb{E}\left[\exp\left(\beta \sum_{s=1} (\widetilde{\lambda}_s - \lambda_s)\right)\right] \cdot \exp(-\beta\varepsilon) \leq \exp(-\beta\varepsilon)$$

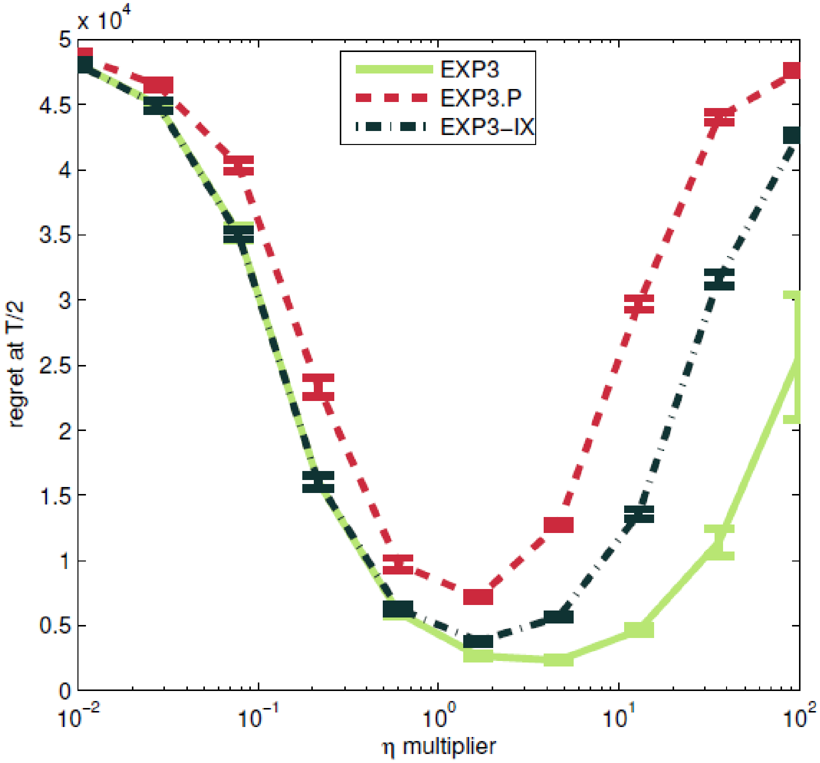
holds for any $\varepsilon > 0$. The statement of the lemma follows from solving for ε.

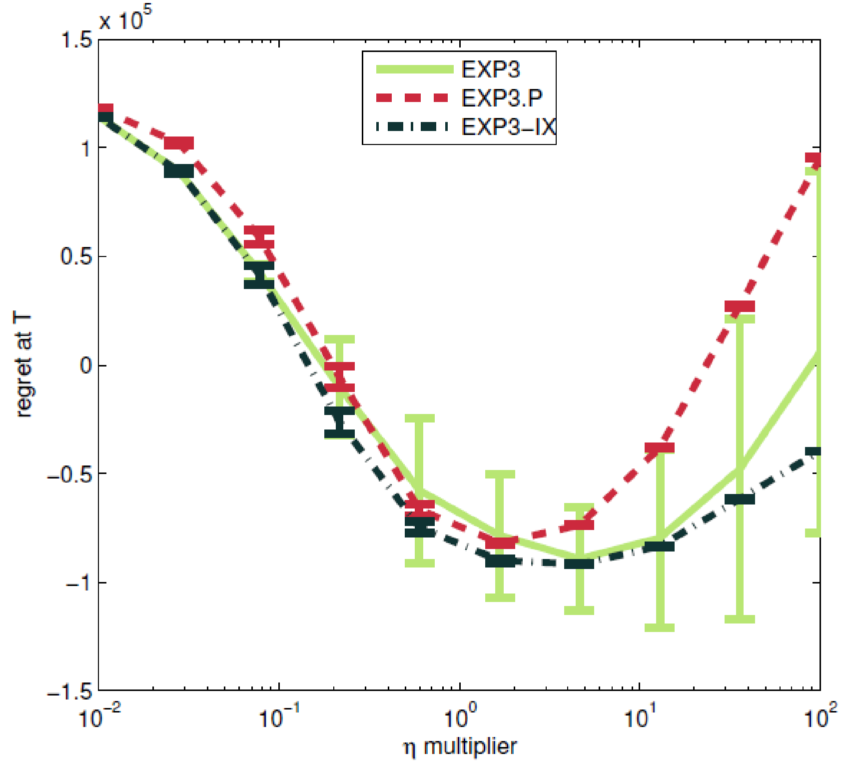
Other results

Setting	Best known HP bound	Our bound
Multi-armed bandits	$5.25\sqrt{KT\log(K/\delta)}$ (Bubeck and Cesa-Bianchi, 2011)	$2\sqrt{2KT\log(K/\delta)}$
Bandits with expert advice	$6\sqrt{KT\log(N/\delta)}$ (Beygelzimer et al., 2011)	$2\sqrt{2KT\log(N/\delta)}$
Tracking the best arm	$7\sqrt{KTS\log(KT/\delta S)}$ (Audibert and Bubeck, 2009)	$2\sqrt{2KTS\log(KT/\delta S)}$
Bandits with side-observations	$\tilde{O}(\sqrt{mT})$ (Alon et al., 2014)	$\tilde{O}(\sqrt{\alpha T})$ $(\alpha \ll m)$

+ whatever you can think about!

Empirical performance





IID losses

non-IID losses

Conclusion

- Implicit exploration is cool
  - Works with losses, no exploration
  - See also Kocák et al. (NIPS 2014) and Neu (COLT 2015)
- Still not well-understood!
  - Tikhonov regularization? Laplace smoothing? Shrinkage?
- Linear bandits?
- Reinforcement learning?
- Active learning?

Conclusion

- Implicit exploration is cool
  - Works with losses, no exploration
  - See also Kocák et al. (NIPS 2014) and Neu (COLT 2015)
- Still not well-understood!
  - Tikhonov regularization? Laplace smoothing? Shrinkage?
- Linear bandits?
- Reinforcement learning?
- Active learning?

THANKS!!!

arXiv:1506.03271