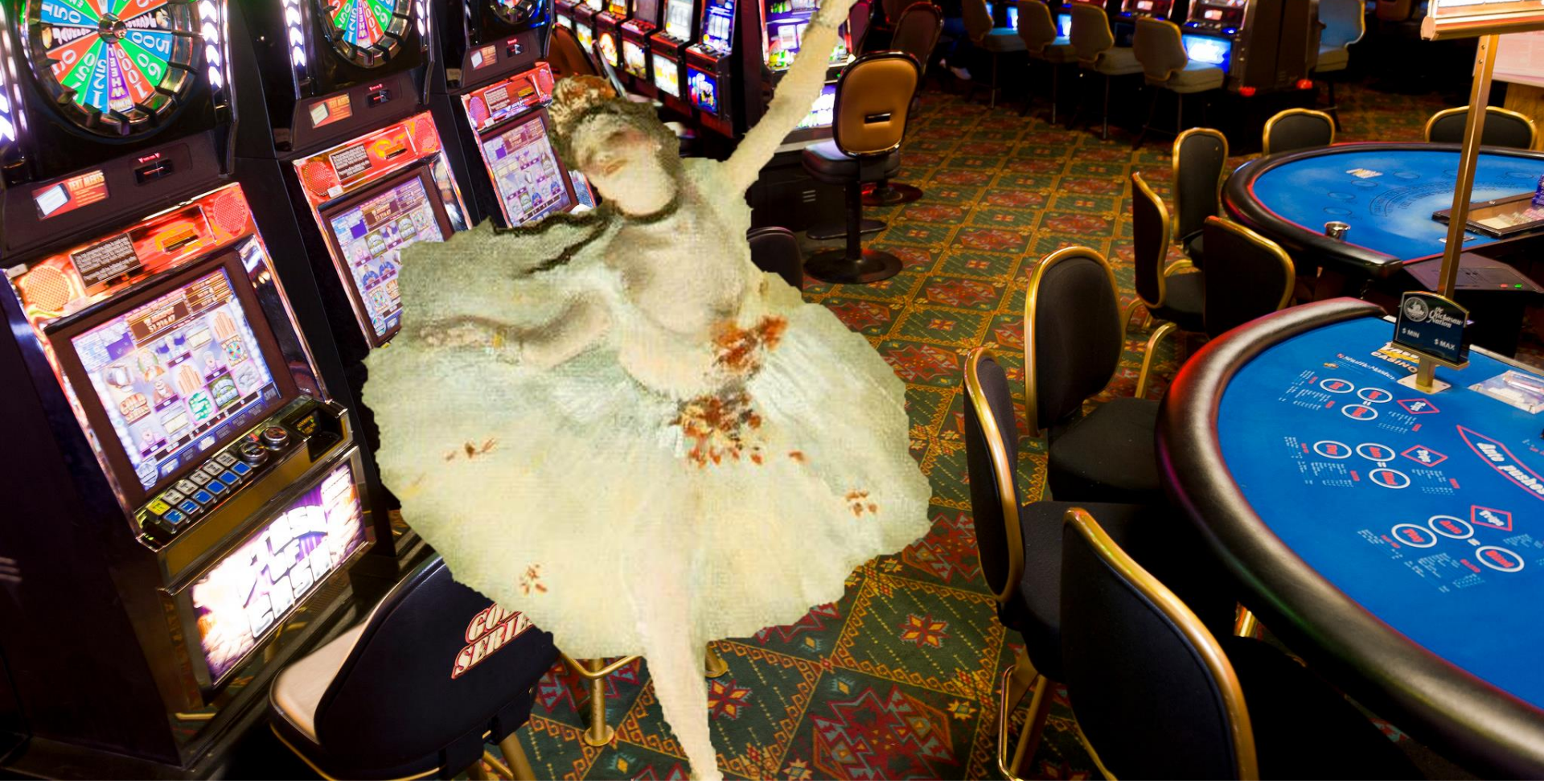
EASINESS IN BANDITS

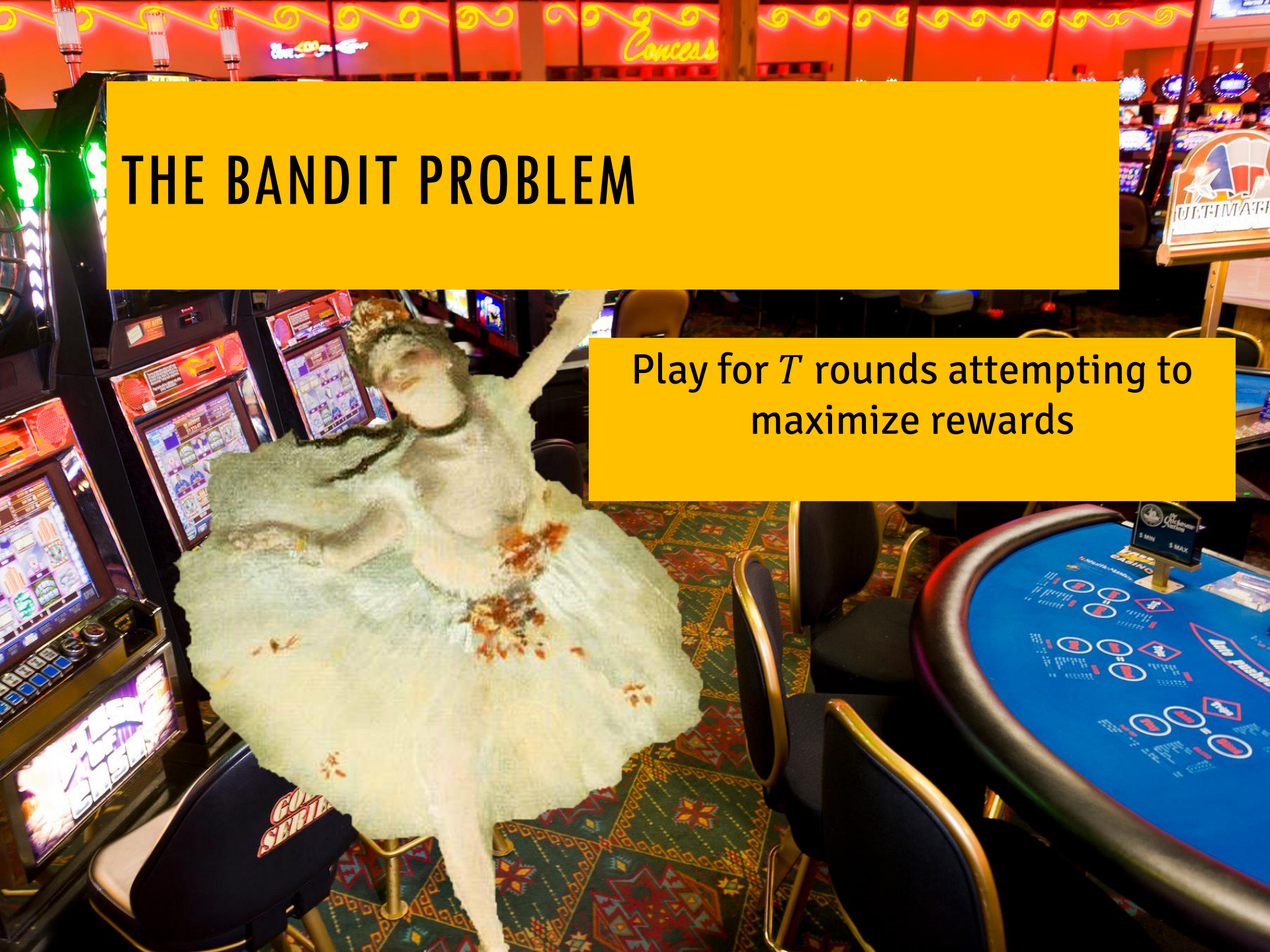
Gergely Neu
Pompeu Fabra University



EASINESS IN BANDITS

Gergely Neu
Pompeu Fabra University

THE BANDIT PROBLEM



Play for T rounds attempting to maximize rewards

THE BANDIT PROBLEM

Play for T rounds attempting to
~~maximize rewards~~
minimize losses

THE BANDIT PROBLEM

Play for T rounds attempting to
~~maximize rewards~~
minimize losses

Need to balance
exploration and exploitation

Motivation:
advertising, clinical trials, …

EASINESS IN BANDITS – A TUTORIAL

Hardness in bandits

- Worst-case upper & lower bounds

Easiness in bandits

- Higher order bounds
- Stochastic bandits and the best of both worlds
- Prior-dependent bounds

NON-STOCHASTIC BANDITS

Parameters:
number of arms K, number of rounds T
Interaction:
For each round $t = 1,2, \dots, T$
- Learner chooses action $I_t \in [K]$
- Environment chooses losses $\ell_{t,i} \in [0,1]$ for all i
- Learner incurs and observes loss ℓ_{t,I_t}

NON-STOCHASTIC BANDITS

Parameters:
number of arms K, number of rounds T
Interaction:
For each round $t = 1,2, \dots, T$

- Learner chooses action $I_t \in [K]$
- Environment chooses losses $\ell_{t,i} \in [0,1]$ for all i
- Learner incurs and observes loss ℓ_{t,I_t}

Goal: minimize **expected** regret

$$\hat{R}_T = \sum_{t=1}^{T} \ell_{t,I_t} - \min_{i \in [K]} \sum_{t=1}^{T} \ell_{t,i}$$

NON-STOCHASTIC BANDITS: LOWER BOUNDS

Theorem (Auer, Cesa-Bianchi, Freund and Schapire, 2002): In the worst case, any algorithm will suffer a regret of $\Omega\left(\sqrt{KT}\right)$

This result also holds for stochastic bandits, as the counterexample is stochastic

NON-STOCHASTIC BANDITS: LOWER BOUNDS

Theorem (Auer, Cesa-Bianchi, Freund and Schapire, 2002): In the worst case, any algorithm will suffer a regret of $\Omega\left(\sqrt{KT}\right)$

This result also holds for stochastic bandits, as the counterexample is stochastic

This talk:
how to go beyond this

NON-STOCHASTIC BANDITS: UPPER BOUNDS

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

Parameter: $\eta > 0$.

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

For each round $t = 1,2, \dots, T$

- For all i, let

$$p_{t,i} = \frac{w_{t,i}}{\sum_j w_{t,j}}.$$

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

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

- For all i, update weight as

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

THE REGRET OF EXP3

Theorem (Auer, Cesa-Bianchi, Freund and Schapire, 2002):
The regret of EXP3 satisfies

$$\hat{R}_T \leq \sqrt{2KT \log K}$$

THE REGRET OF EXP3

Theorem (Auer, Cesa-Bianchi, Freund and Schapire, 2002):
The regret of EXP3 satisfies

$$\hat{R}_T \leq \sqrt{2KT \log K}$$

"Proof":

$$\hat{R}_T \leq \frac{\log K}{\eta} + \frac{\eta}{2}\mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\hat{\ell}_{t,i}^2\right]$$

$$= \frac{\log K}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{i=1}^{K} \ell_{t,i}^2 \leq \frac{\log K}{\eta} + \frac{\eta KT}{2}$$

THE REGRET OF EXP3

Theorem (Auer, Cesa-Bianchi, Freund and Schapire, 2002):
The regret of EXP3 satisfies

$$\hat{R}_T \leq \sqrt{2KT \log K}$$

"Proof":

$$\hat{R}_T \leq \frac{\log K}{\eta} + \frac{\eta}{2}\mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\hat{\ell}_{t,i}^2\right]$$

$$= \frac{\log K}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{i=1}^{K} \ell_{t,i}^2 \leq \frac{\log K}{\eta} + \frac{\eta KT}{2}$$

$$\eta = \sqrt{\frac{2\log K}{KT}}$$

HEY, BUT THAT'S NOT MINIMAX!

Exp3 is strictly suboptimal: you can't remove the $\sqrt{\log K}$ (Audibert, Bubeck and Lugosi, 2014)

HEY, BUT THAT'S NOT MINIMAX!

Exp3 is strictly suboptimal: you can't remove the $\sqrt{\log K}$ (Audibert, Bubeck and Lugosi, 2014)

A minimax algorithm: **PolyINF**

$$p_t = \arg\min_{p \in \Delta_K} \left(\eta p^\top \hat{L}_{t-1} + S_\alpha(p) \right)$$

where $S_\alpha(p)$ is the Tsallis entropy:

$$S_\alpha(p) = \frac{1}{1-\alpha}\left(1 - \sum_{i=1}^{K} p^\alpha\right)$$

HEY, BUT THAT'S NOT MINIMAX!

Exp3 is strictly suboptimal: you can't remove the $\sqrt{\log K}$ (Audibert, Bubeck and Lugosi, 2014)

A minimax algorithm: **PolyINF**

$$p_t = \arg\min_{p \in \Delta_K} \left(\eta p^\top \hat{L}_{t-1} + S_\alpha(p) \right)$$

where $S_\alpha(p)$ is the Tsallis entropy:

$$S_\alpha(p) = \frac{1}{1-\alpha}\left(1 - \sum_{i=1}^K p^\alpha\right)$$

Theorem (Audibert and Bubeck, 2009, Audibert, Bubeck and Lugosi, 2014, Abernethy, Lee and Tewari, 2015):

The regret of PolyINF satisfies $\hat{R}_T \leq 2\sqrt{KT}$

BEYOND MINIMAX #1: HIGHER-ORDER BOUNDS

HIGHER-ORDER BOUNDS

	Full information	Bandit
minimax	$R_T = O\left(\sqrt{T \log K}\right)$	$R_T = O\left(\sqrt{KT}\right)$
first-order $L_{T,i} = \sum_t \ell_{t,i}$	$R_T = O\left(\sqrt{L_{T,i^*} \log K}\right)$	
second-order $S_{T,i} = \sum_t \ell_{t,i}^2$	$R_T = O\left(\sqrt{S_{t,i^*} \log K}\right)$ ★ Cesa-Bianchi, Mansour, Stoltz (2005)	
variance $V_{T,i} = \sum_t \left(\ell_{t,i} - m\right)^2$	$R_T = O\left(\sqrt{V_{T,i^*} \log K}\right)$ ★ Hazan and Kale (2010)	

★ with a little cheating

HIGHER-ORDER BOUNDS

	Full information	Bandit
minimax	$R_T = O\left(\sqrt{T \log K}\right)$	$R_T = O\left(\sqrt{KT}\right)$
first-order $L_{T,i} = \sum_t \ell_{t,i}$	$R_T = O\left(\sqrt{L_{T,i^*} \log K}\right)$	
second-order $S_{T,i} = \sum_t \ell_{t,i}^2$	$R_T = O\left(\sqrt{S_{t,i^*} \log K}\right)$ ★ Cesa-Bianchi, Mansour, Stoltz (2005)	Easy!
variance $V_{T,i} = \sum_t \left(\ell_{t,i} - m\right)^2$	$R_T = O\left(\sqrt{V_{T,i^*} \log K}\right)$ ★ Hazan and Kale (2010)	

★ with a little cheating

SECOND-ORDER BOUNDS

Easy!

The Exp3 “proof”:

$$\hat{R}_T \leq \frac{\log K}{\eta} + \frac{\eta}{2}\mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\hat{\ell}_{t,i}^2\right]$$

$$= \frac{\log K}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{i=1}^{K} \ell_{t,i}^2 \leq \frac{\log K}{\eta} + \frac{\eta K T}{2}$$

SECOND-ORDER BOUNDS

Easy!

The Exp3 "proof":

$$\hat{R}_T \leq \frac{\log K}{\eta} + \frac{\eta}{2}\mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\hat{\ell}_{t,i}^2\right]$$

$$= \frac{\log K}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{i=1}^{K} \ell_{t,i}^2 \leq \frac{\log K}{\eta} + \frac{\eta KT}{2}$$

SECOND-ORDER BOUNDS

Easy!

The Exp3 "proof":

$$\hat{R}_T \leq \frac{\log K}{\eta} + \frac{\eta}{2}\mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\hat{\ell}_{t,i}^2\right]$$

$$= \frac{\log K}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{i=1}^{K} \ell_{t,i}^2 \leq \frac{\log K}{\eta} + \frac{\eta KT}{2}$$

Theorem (Auer, Cesa-Bianchi, Freund and Schapire, 2002):
The regret of EXP3 satisfies

$$\hat{R}_T \leq \sqrt{2\log K \sum_{t=1}^{T}\sum_{i=1}^{K} \ell_{t,i}^2}$$

HIGHER-ORDER BOUNDS

	Full information	Bandit
minimax	$R_T = O\left(\sqrt{T \log K}\right)$	$R_T = O\left(\sqrt{KT}\right)$
first-order $L_{T,i} = \sum_t \ell_{t,i}$	$R_T = O\left(\sqrt{L_{T,i^*} \log K}\right)$	
second-order $S_{T,i} = \sum_t \ell_{t,i}^2$	$R_T = O\left(\sqrt{S_{t,i^*} \log K}\right)$ ★ Cesa-Bianchi, Mansour, Stoltz (2005)	Easy!
variance $V_{T,i} = \sum_t \left(\ell_{t,i} - m\right)^2$	$R_T = O\left(\sqrt{V_{T,i^*} \log K}\right)$ ★ Hazan and Kale (2010)	

★ with a little cheating

HIGHER-ORDER BOUNDS

	Full information	Bandit
minimax	$R_T = O\left(\sqrt{T \log K}\right)$	$R_T = O\left(\sqrt{KT}\right)$
first-order $L_{T,i} = \sum_t \ell_{t,i}$	$R_T = O\left(\sqrt{L_{T,i^*} \log K}\right)$	
second-order $S_{T,i} = \sum_t \ell_{t,i}^2$	$R_T = O\left(\sqrt{S_{t,i^*} \log K}\right)$ ★ Cesa-Bianchi, Mansour, Stoltz (2005)	$R_T = \tilde{O}\left(\sqrt{\sum_i S_{t,i}}\right)$ Auer et al. (2002) + some hacking
variance $V_{T,i} = \sum_t \left(\ell_{t,i} - m\right)^2$	$R_T = O\left(\sqrt{V_{T,i^*} \log K}\right)$ ★ Hazan and Kale (2010)	

★ with a little cheating

HIGHER-ORDER BOUNDS

	Full information	Bandit
minimax	$R_T = O\left(\sqrt{T \log K}\right)$	$R_T = O\left(\sqrt{KT}\right)$
first-order $L_{T,i} = \sum_t \ell_{t,i}$	$R_T = O\left(\sqrt{L_{T,i^*} \log K}\right)$	
second-order $S_{T,i} = \sum_t \ell_{t,i}^2$	$R_T = O\left(\sqrt{S_{t,i^*} \log K}\right)$ ★ Cesa-Bianchi, Mansour, Stoltz (2005)	$R_T = \tilde{O}\left(\sqrt{\sum_i S_{t,i}}\right)$ Auer et al. (2002) + some hacking
variance $V_{T,i} = \sum_t \left(\ell_{t,i} - m\right)^2$	$R_T = O\left(\sqrt{V_{T,i^*} \log K}\right)$ ★ Hazan and Kale (2010)	not so easy

★ with a little cheating

VARIANCE BOUNDS

not so easy

Need to replace $\sum_t \sum_i \ell_{t,i}^2$ by $\sum_t \sum_i \left(\ell_{t,i} - \mu_{T,i}\right)^2$,
where $\mu_{T,i} = \frac{1}{T}\sum_t \ell_{t,i}$

VARIANCE BOUNDS

not so easy

Need to replace $\sum_t \sum_i \ell_{t,i}^2$ by $\sum_t \sum_i \left(\ell_{t,i} - \mu_{T,i}\right)^2$, where $\mu_{T,i} = \frac{1}{T}\sum_t \ell_{t,i}$

Hazan and Kale (2011), heavily paraphrased:

- Replace $\mu_{T,i}$ by $\mu_{t,i}$ (easy)
- Estimate $\mu_{t,i}$ by an appropriate $\tilde{\mu}_{t,i}$: reservoir sampling in exploration rounds
- Use **Exp3** with loss estimates

$$\hat{\ell}_{t,i} = \frac{\ell_{t,i} - \tilde{\mu}_{t,i}}{p_{t,i}} + \tilde{\mu}_{t,i}$$

VARIANCE BOUNDS

not so easy

Need to replace $\sum_t \sum_i \ell_{t,i}^2$ by $\sum_t \sum_i \left(\ell_{t,i} - \mu_{T,i}\right)^2$,
where $\mu_{T,i} = \frac{1}{T}\sum_t \ell_{t,i}$

Hazan and Ka

- Replace $\mu_{T,i}$ b
- Estimate $\mu_{t,i}$ b ir sampling in ex

But that doesn't work!

- Use **Exp3** with loss estimates

$$\hat{\ell}_{t,i} = \frac{\ell_{t,i} - \tilde{\mu}_{t,i}}{p_{t,i}} + \tilde{\mu}_{t,i}$$

THE RIGHT WAY TO GET VARIANCE BOUNDS

Instead of **Exp3**, use SCRiBLe:

$$p_t = \arg\min_{p \in \Delta_K} \left(p^\top \hat{L}_{t-1} + \Psi(p) \right)$$

with $\hat{L}_{t-1,i} = \sum_{s=1}^{t-1} \left(\hat{c}_{t,i} + \tilde{\mu}_{t,i} \right)$

THE RIGHT WAY TO GET VARIANCE BOUNDS

"self-concordant regularizer"

Instead of **Exp3**, use SCRiBLe:

$$p_t = \arg\min_{p \in \Delta_K} \left(p^\top \hat{L}_{t-1} + \Psi(p) \right)$$

with $\hat{L}_{t-1,i} = \sum_{s=1}^{t-1} \left(\hat{c}_{t,i} + \tilde{\mu}_{t,i} \right)$

THE RIGHT WAY TO GET VARIANCE BOUNDS

"self-concordant regularizer"

Instead of **Exp3**, use SCRiBLe:

$$p_t = \arg\min_{p \in \Delta_K} \left(p^\top \hat{L}_{t-1} + \Psi(p) \right)$$

with $\hat{L}_{t-1,i} = \sum_{s=1}^{t-1} \left(\hat{c}_{t,i} + \tilde{\mu}_{t,i} \right)$

$\hat{c}_{t,i}$ ≈appropriate unbiased estimate of $\ell_{t,i} - \tilde{\mu}_{t,i}$

THE RIGHT WAY TO GET VARIANCE BOUNDS

"self-concordant regularizer"

Instead of **Exp3**, use SCRiBLe:

$$p_t = \arg\min_{p \in \Delta_K} \left(p^\top \hat{L}_{t-1} + \Psi(p) \right)$$

with $\hat{L}_{t-1,i} = \sum_{s=1}^{t-1} \left(\hat{c}_{t,i} + \tilde{\mu}_{t,i} \right)$

$\hat{c}_{t,i}$ ≈appropriate unbiased

Theorem (Hazan and Kale, 2011):
The regret of the above algorithm satisfies

$$\hat{R}_T = \tilde{O}\left(K^2 \sqrt{\sum_{t=1}^{T} \sum_{i=1}^{K} \left(\ell_{t,i}^2 - \mu_{T,i} \right)^2} \right)$$

HIGHER-ORDER BOUNDS

	Full information	Bandit
minimax	$R_T = O\left(\sqrt{T \log K}\right)$	$R_T = O\left(\sqrt{KT}\right)$
first-order $L_{T,i} = \sum_t \ell_{t,i}$	$R_T = O\left(\sqrt{L_{T,i^*} \log K}\right)$	
second-order $S_{T,i} = \sum_t \ell_{t,i}^2$	$R_T = O\left(\sqrt{S_{t,i^*} \log K}\right)$ ★ Cesa-Bianchi, Mansour, Stoltz (2005)	$R_T = \tilde{O}\left(\sqrt{\sum_i S_{t,i}}\right)$ Auer et al. (2002) + some hacking
variance $V_{T,i} = \sum_t \left(\ell_{t,i} - m\right)^2$	$R_T = O\left(\sqrt{V_{T,i^*} \log K}\right)$ ★ Hazan and Kale (2010)	not so easy

★ with a little cheating

HIGHER-ORDER BOUNDS

	Full information	Bandit
minimax	$R_T = O\left(\sqrt{T \log K}\right)$	$R_T = O\left(\sqrt{KT}\right)$
first-order $L_{T,i} = \sum_t \ell_{t,i}$	$R_T = O\left(\sqrt{L_{T,i^*} \log K}\right)$	
second-order $S_{T,i} = \sum_t \ell_{t,i}^2$	$R_T = O\left(\sqrt{S_{t,i^*} \log K}\right)$ ★ Cesa-Bianchi, Mansour, Stoltz (2005)	$R_T = \tilde{O}\left(\sqrt{\sum_i S_{t,i}}\right)$ Auer et al. (2002) + some hacking
variance $V_{T,i} = \sum_t \left(\ell_{t,i} - m\right)^2$	$R_T = O\left(\sqrt{V_{T,i^*} \log K}\right)$ ★ Hazan and Kale (2010)	$R_T = \tilde{O}\left(K^2 \sqrt{\sum_i V_{t,i}}\right)$ Hazan and Kale (2011)

★ with a little cheating

HIGHER-ORDER BOUNDS

	Full information	Bandit
minimax	$R_T = O\left(\sqrt{T \log K}\right)$	$R_T = O\left(\sqrt{KT}\right)$
first-order $L_{T,i} = \sum_t \ell_{t,i}$	$R_T = O\left(\sqrt{L_{T,i^*} \log K}\right)$	should be easy?
second-order $S_{T,i} = \sum_t \ell_{t,i}^2$	$R_T = O\left(\sqrt{S_{t,i^*} \log K}\right)$ ★ Cesa-Bianchi, Mansour, Stoltz (2005)	$R_T = \tilde{O}\left(\sqrt{\sum_i S_{t,i}}\right)$ Auer et al. (2002) + some hacking
variance $V_{T,i} = \sum_t \left(\ell_{t,i} - m\right)^2$	$R_T = O\left(\sqrt{V_{T,i^*} \log K}\right)$ ★ Hazan and Kale (2010)	$R_T = \tilde{O}\left(K^2 \sqrt{\sum_i V_{t,i}}\right)$ Hazan and Kale (2011)

★ with a little cheating

FIRST-ORDER BOUNDS

should be easy?

"Small-gain" bounds:

- Consider the gain game with $g_{t,i} = 1 - \ell_{t,i}$
- Auer, Cesa-Bianchi, Freund and Schapire (2002):

$$R_T = O\left(\sqrt{K G_{T,i^*} \log K}\right)$$

$G_{T,i} = \sum_t g_{t,i}$

FIRST-ORDER BOUNDS

should be easy?

"Small-gain" bounds:

- Consider the gain game with $g_{t,i} = 1 - \ell_{t,i}$
- Auer, Cesa-Bianchi, Freund and Schapire (2002):

$$R_T = O\left(\sqrt{K G_{T,i^*} \log K}\right)$$

$G_{T,i} = \sum_t g_{t,i}$

Problem:
only good if best expert is bad!

FIRST-ORDER BOUNDS

should be easy?

"Small-gain" bounds: $R_T = O\left(\sqrt{K G_{T,i^*} \log K}\right)$

A little trickier analysis gives

$R_T = O\left(\sqrt{\sum_t \sum_i g_{t,i} \log K}\right)$ or $R_T = O\left(\sqrt{\sum_t \sum_i \ell_{t,i} \log K}\right)$

FIRST-ORDER BOUNDS

should be easy?

"Small-gain" bounds: $R_T = O\left(\sqrt{K G_{T,i^*} \log K}\right)$

A little trickier analysis gives

$R_T = O\left(\sqrt{\sum_t \sum_i g_{t,i} \log K}\right)$ or $R_T = O\left(\sqrt{\sum_t \sum_i \ell_{t,i} \log K}\right)$

Problem:
one misbehaving action ruins the bound!

FIRST-ORDER BOUNDS

should be easy?

"Small-gain" bounds: $R_T = O\left(\sqrt{K G_{T,i^*} \log K}\right)$

A little trickier analysis gives $R_T = O\left(\sqrt{\sum_t \sum_i \ell_{t,i} \log K}\right)$

Actual first-order bounds:

- Stoltz (2005): $K\sqrt{L_T^*}$
- Allenberg, Auer, Györfi and Ottucsák (2006): $\sqrt{K L_T^*}$
- Rakhlin and Sridharan (2013): $K^{3/2}\sqrt{L_T^*}$

FIRST-ORDER BOUNDS

should be easy?

"Small-gain" bounds: $R_T = O\left(\sqrt{K G_{T,i^*} \log K}\right)$

A little trickier analysis gives $R_T = O\left(\sqrt{\sum_t \sum_i \ell_{t,i} \log K}\right)$

Actual first-order bounds:

- Stoltz (2005): $K\sqrt{L_T^*}$
- Allenberg, Auer, Györfi and Ottucsák (2006): $\sqrt{K L_T^*}$
- Rakhlin and Sridharan (2013): $K^{3/2}\sqrt{L_T^*}$

THE GREEN ALGORITHM (ALLENBERG ET AL., 2006)

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

Parameter: $\eta > 0$.

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

For each round $t = 1,2, \dots, T$

- For all i, let

$$p_{t,i} = \frac{w_{t,i}}{\sum_j w_{t,j}}.$$

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

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

- For all i, update weight as

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

THE GREEN ALGORITHM (ALLENBERG ET AL., 2006)

Green (Allenberg, Auer, Györfi and Ottucsák, 2006)

Parameters: $\eta > 0, \gamma \in (0,1)$.

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

For each round $t = 1,2, \dots, T$

- For all i, let

$$p_{t,i} = \frac{w_{t,i}}{\sum_j w_{t,j}} \text{ and let } \tilde{p}_{t,i} = 0 \text{ if } p_{t,i} \leq \gamma.$$

- Draw $I_t \sim \tilde{\boldsymbol{p}}_t$.
- For all i, let

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

- For all i, update weight as

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

THE GREEN ALGORITHM (ALLENBERG ET AL., 2006)

Analysis idea:

- As long as $p_{t,i} \geq \gamma$ for an i, we have

$$\hat{L}_{t-1,i} \leq \hat{L}_{t-1,j} + \tilde{O}(\log(1/\gamma)\,/\eta)$$

THE GREEN ALGORITHM (ALLENBERG ET AL., 2006)

Analysis idea:

- As long as $p_{t,i} \geq \gamma$ for an i, we have

$$\hat{L}_{t-1,i} \leq \hat{L}_{t-1,j} + \tilde{O}(\log(1/\gamma)\,/\eta)$$

"the loss estimates are not too far apart"

THE GREEN ALGORITHM (ALLENBERG ET AL., 2006)

Analysis idea:

- As long as $p_{t,i} \geq \gamma$ for an i, we have

$$\hat{L}_{t-1,i} \leq \hat{L}_{t-1,j} + \tilde{O}(\log(1/\gamma)/\eta)$$

"the loss estimates are not too far apart"

- Once $p_{t,i} \leq \gamma$ occurs, $\hat{L}_{t,i}$ stops growing, so

$$\hat{L}_{T,i} \leq \hat{L}_{T,j} + \tilde{O}(\log(1/\gamma)/\eta) + \tilde{O}(1/\gamma)$$

THE GREEN ALGORITHM (ALLENBERG ET AL., 2006)

Getting back to the **Exp3** proof:

$$\hat{R}_T \leq \frac{\log K}{\eta} + \frac{\eta}{2}\mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\hat{\ell}_{t,i}^2\right]$$

THE GREEN ALGORITHM (ALLENBERG ET AL., 2006)

Getting back to the **Exp3** proof:

$$\hat{R}_T \leq \frac{\log K}{\eta} + \frac{\eta}{2}\mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\hat{\ell}_{t,i}^2\right]$$

$$\leq \frac{\log K}{\eta} + \frac{\eta}{2}\mathbf{E}\left[\sum_{i=1}^{K} \hat{L}_{T,i}\right]$$

THE GREEN ALGORITHM (ALLENBERG ET AL., 2006)

Getting back to the **Exp3** proof:

$$\begin{aligned}\hat{R}_T &\leq \frac{\log K}{\eta} + \frac{\eta}{2}\mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\hat{\ell}_{t,i}^2\right] \\ &\leq \frac{\log K}{\eta} + \frac{\eta}{2}\mathbf{E}\left[\sum_{i=1}^{K}\hat{L}_{T,i}\right] \\ &\leq \frac{\log K}{\eta} + \frac{\eta}{2}\mathbf{E}\left[K\hat{L}_{T,i^*}\right] + \tilde{O}(K)\end{aligned}$$

THE GREEN ALGORITHM (ALLENBERG ET AL., 2006)

Getting back to the **Exp3** proof:

$$\begin{aligned}
\hat{R}_T &\leq \frac{\log K}{\eta} + \frac{\eta}{2}\mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\hat{\ell}_{t,i}^2\right] \\
&\leq \frac{\log K}{\eta} + \frac{\eta}{2}\mathbf{E}\left[\sum_{i=1}^{K} \hat{L}_{T,i}\right] \\
&\leq \frac{\log K}{\eta} + \frac{\eta}{2}\mathbf{E}\left[K\hat{L}_{T,i^*}\right] + \tilde{O}(K) \\
&\leq \frac{\log K}{\eta} + \frac{\eta}{2}KL_{T,i^*} + \tilde{O}(K)
\end{aligned}$$

THE GREEN ALGORITHM (ALLENBERG ET AL., 2006)

Getting back to the **Exp3** proof:

Theorem (Allenberg et al., 2006):
The regret of **Green** satisfies

$$\hat{R}_T = \tilde{O}\left(\sqrt{KL_T^*} + K\right)$$

$$\leq \frac{\log K}{\eta} + \frac{\eta}{2}\mathbf{E}\left[K\hat{L}_{T,i^*}\right] + \tilde{O}(K)$$

$$\leq \frac{\log K}{\eta} + \frac{\eta}{2}KL_{T,i^*} + \tilde{O}(K)$$

A SIMPLER ALGORITHM: EXP3-IX

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

Parameter: $\eta > 0$.

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

For each round $t = 1,2, \dots, T$

- For all i, let

$$p_{t,i} = \frac{w_{t,i}}{\sum_j w_{t,j}}.$$

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

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

- For all i, update weight as

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

A SIMPLER ALGORITHM: EXP3-IX

EXP3-IX (Kocák et al., 2014, Neu 2015a, Neu 2015b)

Parameter: $\eta > 0, \gamma > 0$.

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

For each round $t = 1,2, \dots, T$

- For all i, let

$$p_{t,i} = \frac{w_{t,i}}{\sum_j w_{t,j}}.$$

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

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

- For all i, update weight as

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

A SIMPLER ALGORITHM: EXP3-IX

EXP3-IX (Kocák et al., 2014, Neu 2015a, Neu 2015b)

Theorem (Neu, 2015):
The regret of **Exp3-IX** satisfies

$$\hat{R}_T = \tilde{O}\left(\sqrt{KL_T^*} + K\right)$$

$$p_{t,i} = \frac{w_{t,i}}{\sum_j w_{t,j}}.$$

- Draw $I_t \sim \boldsymbol{p}_t$.
- 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 weight as

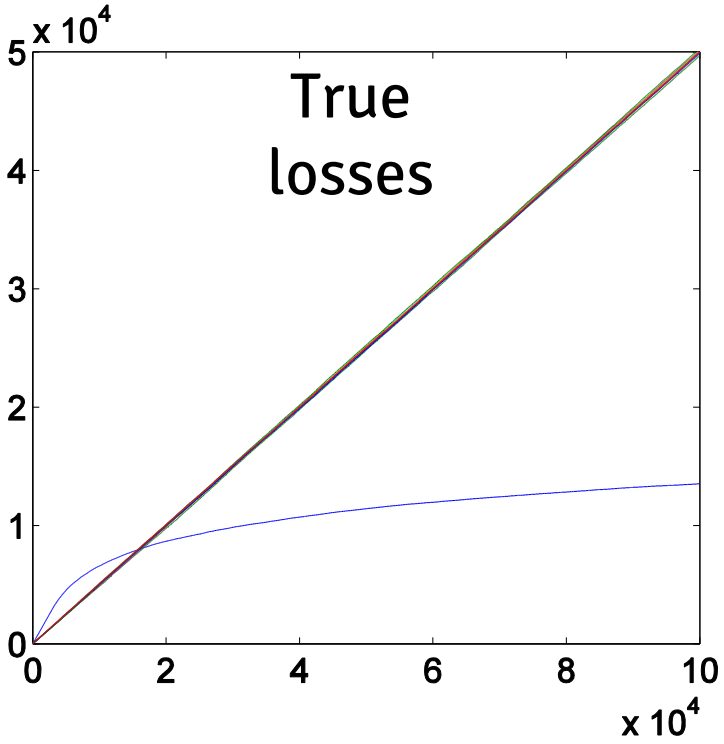
$$w_{t+1,i} = w_{t,i} e^{-\eta \hat{\ell}_{t,i}}$$

IMPLICIT EXPLORATION IN ACTION

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

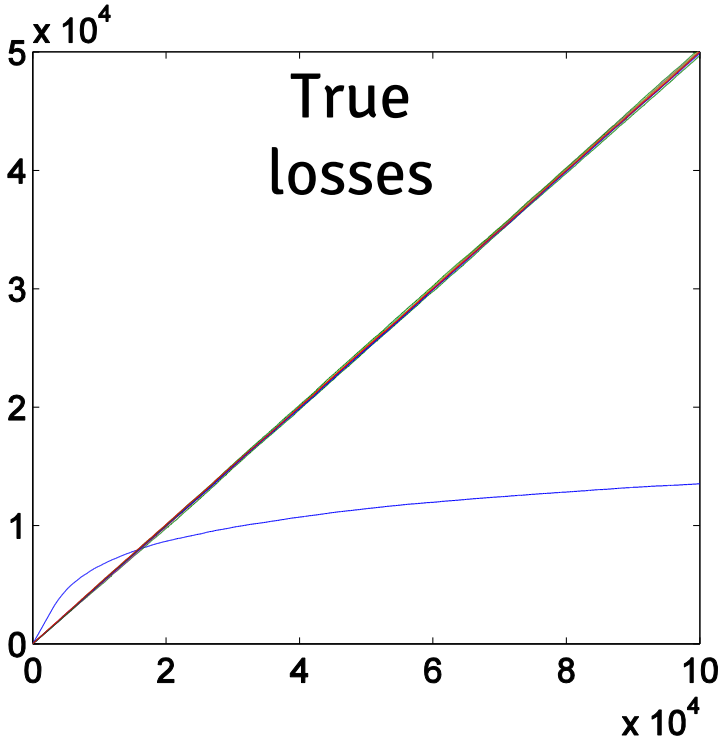
IMPLICIT EXPLORATION IN ACTION

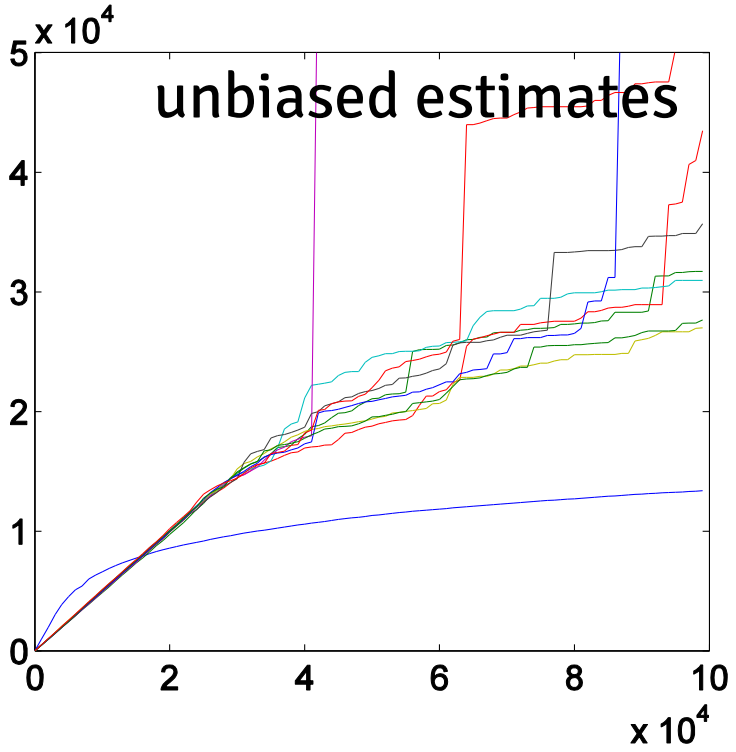
$$\widehat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i} + \gamma} \mathbf{1}_{\{I_t = i\}}$$



IMPLICIT EXPLORATION IN ACTION

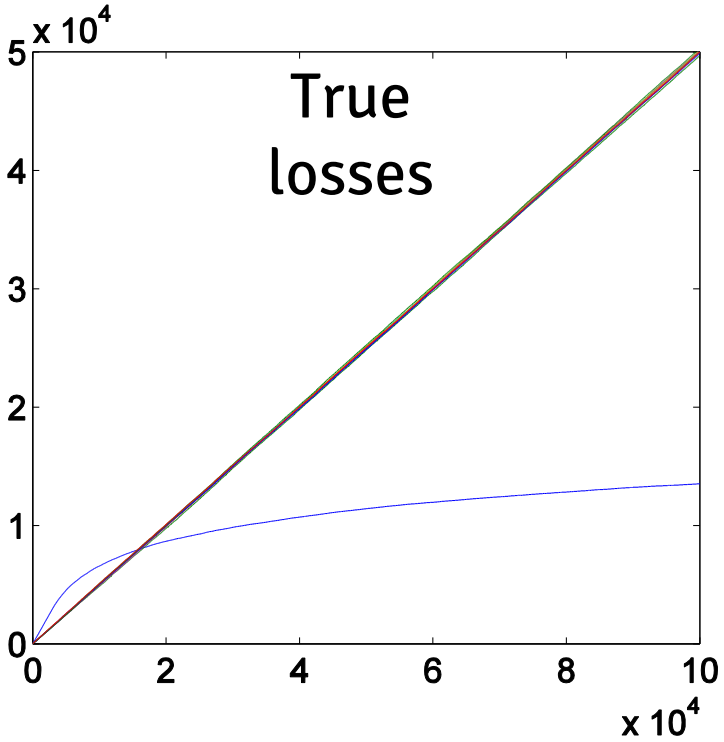
$$\widehat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i} + \gamma} \mathbf{1}_{\{I_t = i\}}$$

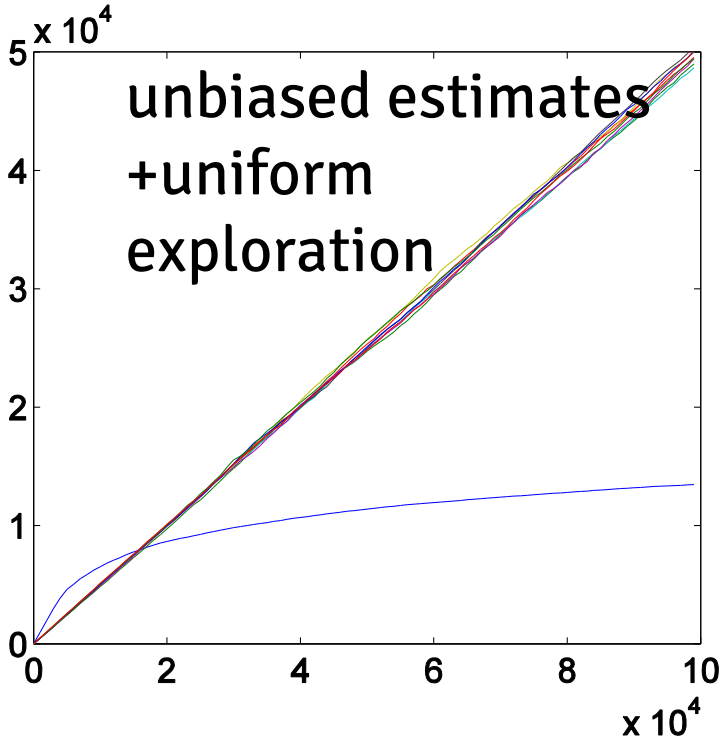




IMPLICIT EXPLORATION IN ACTION

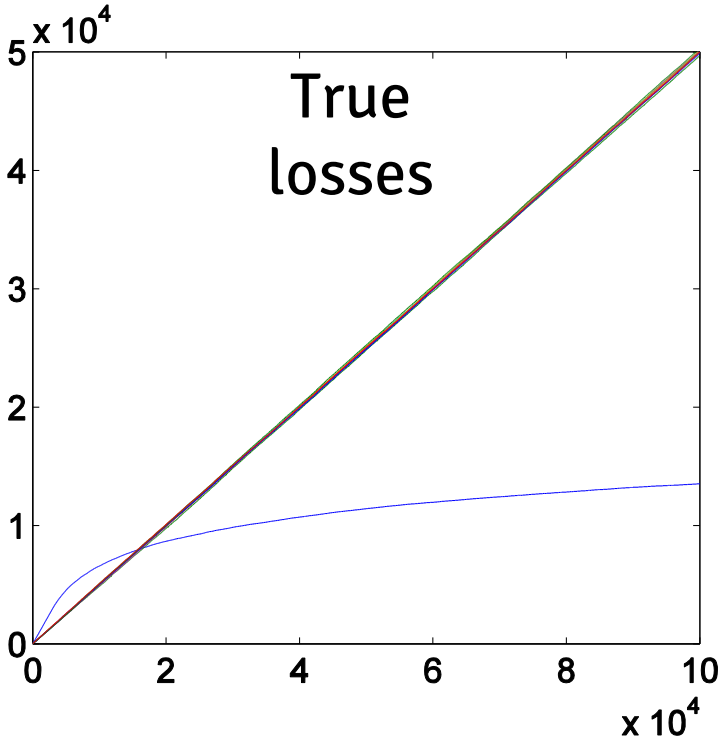
$$\widehat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i} + \gamma} \mathbf{1}_{\{I_t = i\}}$$

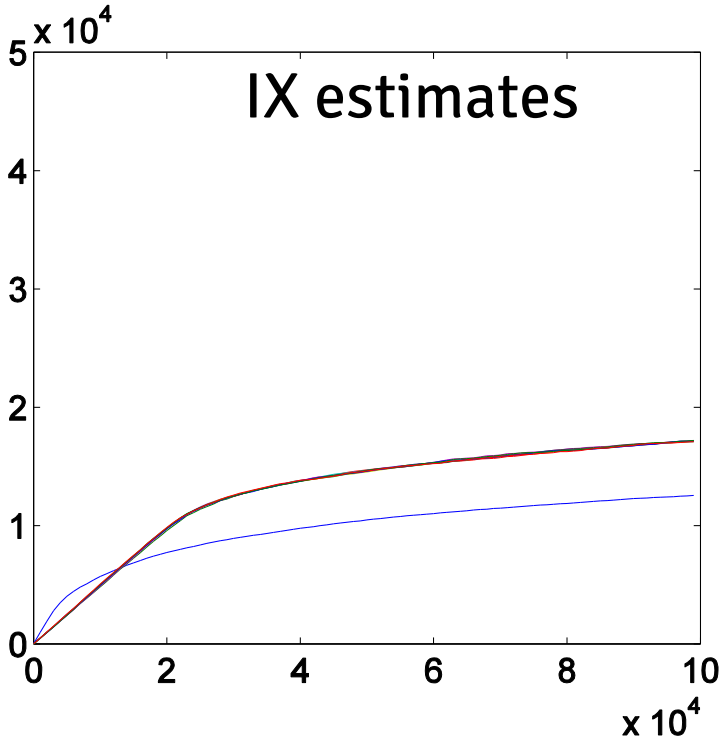




IMPLICIT EXPLORATION IN ACTION

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





HIGHER-ORDER BOUNDS

	Full information	Bandit
minimax	$R_T = O\left(\sqrt{T \log K}\right)$	$R_T = O\left(\sqrt{KT}\right)$
first-order $L_{T,i} = \sum_t \ell_{t,i}$	$R_T = O\left(\sqrt{L_{T,i^*} \log K}\right)$	should be easy?
second-order $S_{T,i} = \sum_t \ell_{t,i}^2$	$R_T = O\left(\sqrt{S_{t,i^*} \log K}\right)$ ★ Cesa-Bianchi, Mansour, Stoltz (2005)	$R_T = \tilde{O}\left(\sqrt{\sum_i S_{t,i}}\right)$ Auer et al. (2002) + some hacking
variance $V_{T,i} = \sum_t \left(\ell_{t,i} - m\right)^2$	$R_T = O\left(\sqrt{V_{T,i^*} \log K}\right)$ ★ Hazan and Kale (2010)	$R_T = \tilde{O}\left(K^2 \sqrt{\sum_i V_{t,i}}\right)$ Hazan and Kale (2011)

★ with a little cheating

HIGHER-ORDER BOUNDS

	Full information	Bandit
minimax	$R_T = O\left(\sqrt{T \log K}\right)$	$R_T = O\left(\sqrt{KT}\right)$
first-order $L_{T,i} = \sum_t \ell_{t,i}$	$R_T = O\left(\sqrt{L_{T,i^*} \log K}\right)$	$R_T = \tilde{O}\left(\sqrt{K L_{T,i^*}}\right)$
second-order $S_{T,i} = \sum_t \ell_{t,i}^2$	$R_T = O\left(\sqrt{S_{t,i^*} \log K}\right)$ ★ Cesa-Bianchi, Mansour, Stoltz (2005)	$R_T = \tilde{O}\left(\sqrt{\sum_i S_{t,i}}\right)$ Auer et al. (2002) + some hacking
variance $V_{T,i} = \sum_t \left(\ell_{t,i} - m\right)^2$	$R_T = O\left(\sqrt{V_{T,i^*} \log K}\right)$ ★ Hazan and Kale (2010)	$R_T = \tilde{O}\left(K^2 \sqrt{\sum_i V_{t,i}}\right)$ Hazan and Kale (2011)

★ with a little cheating

HIGHER-ORDER LOWER BOUNDS

Gerchinovitz and Lattimore (2016), heavily paraphrased:

Theorem:
No algorithm can do better than
$$\hat{R}_T = \Omega\left(\sqrt{L_T^* K}\right)$$

HIGHER-ORDER LOWER BOUNDS

Gerchinovitz and Lattimore (2016), heavily paraphrased:

Theorem:
No algorithm can do better than
$\hat{R}_T = \Omega\left(\sqrt{L_T^* K}\right)$

Theorem:
"No algorithm can do better than
$\hat{R}_T = \Omega\left(\sqrt{\sum_i V_{t,i}}\right)$"

BEYOND MINIMAX #2: STOCHASTIC LOSSES AND THE "BEST OF BOTH WORLDS"

Lai - Robbins '86 - paraphrased

Asymptotics!

$$\mathcal{E} = \{ \nu \mid \nu = (N(\mu_1, 1), \dots, N(\mu_k, 1)), \mu_1, \dots, \mu_k \in \mathbb{R} \}$$

Gaussian environments

$$\Pi = \{ \pi \mid \forall \nu \in \mathcal{E}, p > 0 : R_n(\pi, \nu) = O(n^p) \}$$

"consistent policies"

Instance optimality

① $\forall \pi \in \Pi, \forall \nu \in \mathcal{E}$: $\liminf_{n \to \infty} \frac{R_n(\pi, \nu)}{\log(n)} \geq \sum_{i : \Delta_i(\nu) > 0} \frac{2}{\Delta_i(\nu)} = C^*(\nu)$

② $\exists \pi \in \Pi$ s.t. $\forall \nu \in \mathcal{E}$: $\limsup_{n \to \infty} \frac{R_n(\pi, \nu)}{\log(n)} = C^*(\nu)$.

Lai-Robbins '86 - paraphrased

Asymptotics!

$$\mathcal{E} = \{ \nu \mid \nu = (N(\mu_1, 1), \dots, N(\mu_K, 1)), \mu_1, \dots, \mu_K \in \mathbb{R} \}$$

Gaussian environments

$$\Pi = \{ \pi \mid \forall \nu \in \mathcal{E}, p > 0 : R_n(\pi, \nu) = O(n^p) \}$$

"consistent policies"

Instance optimality

1. $\forall \pi \in \Pi, \forall \nu \in \mathcal{E} : \liminf \frac{R_n(\pi, \nu)}{\log(n)} \geq \sum_{i : \Delta_i(\nu) > 0} \frac{2}{\Delta_i(\nu)} \doteq C^*(\nu)$

2. $\exists \pi \in \Pi$ s.t. $\forall \nu \in \mathcal{E} : \limsup_{n \to \infty} \frac{R_n(\pi, \nu)}{\log(n)} = C^*(\nu).$

Lai-Robbins '86 - paraphrased

Asymptotics!

$$\mathcal{E} = \{ \nu \mid \nu = (N(\mu_1, 1), \dots, N(\mu_K, 1)), \mu_1, \dots, \mu_K \in \mathbb{R} \}$$

Ga

$$\Pi = \{ \dots = O(n^p) \}$$

"co ality

TL;DR:

$$\hat{R}_T = O(C(\nu) \log T)$$

is achievable for i.i.d. losses

1. $\forall \pi \in \Pi, \forall \nu \in \mathcal{E}$: $\liminf \dots \geq \sum_{i:\Delta_i(\nu)>0} \frac{2}{\Delta_i(\nu)} \doteq C^*(\nu)$

2. $\exists \pi \in \Pi$ s.t. $\forall \nu \in \mathcal{E}$: $\limsup_{n\to\infty} \frac{R_n(\pi, \nu)}{\log(n)} = C^*(\nu)$.

THE BEST OF BOTH WORLDS

Is it possible to come up with an algorithm with

$$\hat{R}_T = \tilde{O}\left(\sqrt{KT}\right)$$

for non-stochastic losses and

$$\hat{R}_T = O(C(\nu)\log T)$$

for stochastic losses?

THE BEST OF BOTH WORLDS

Is it possible to come up with an algorithm with

$$\hat{R}_T = \tilde{O}\left(\sqrt{KT}\right)$$

for non-stochastic losses and

$$\hat{R}_T = O(C(\nu)\log T)$$

for stochastic losses?

YES*!!

*almost

THE BEST OF BOTH WORLDS: ALGORITHMS

Bubeck and Slivkins (2012):

- Assume that environment is stochastic, act aggressively
- If the losses fail on a stochasticity test, then fall back to **Exp3**
- **Regret:** $\tilde{O}(\sqrt{KT})$ on adversarial, $O(\log^2 T)$ on stochastic

THE BEST OF BOTH WORLDS: ALGORITHMS

Bubeck and Slivkins (2012):

- Assume that environment is stochastic, act aggressively
- If the losses fail on a stochasticity test, then fall back to **Exp3**
- **Regret:** $\tilde{O}(\sqrt{KT})$ on adversarial, $O(\log^2 T)$ on stochastic

Auer and Chiang (2016), see Peter's talk tomorrow:

- Better test, better algorithm for stochastic losses
- **Regret:** $O(\sqrt{KT \log K})$ on adversarial, $O(\tilde{C}(\nu) \log T)$ on stochastic

A SIMPLE ALGORITHM: EXP3++ (SELDIN AND SLIVKINS, 2014)

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

Parameter: $\eta > 0$.

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

For each round $t = 1,2, \dots, T$

- For all i, let

$$p_{t,i} = \frac{w_{t,i}}{\sum_j w_{t,j}}.$$

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

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

- For all i, update weight as

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

A SIMPLE ALGORITHM: EXP3++ (SELDIN AND SLIVKINS, 2014), PARAPHRASED

EXP3++ (SS, 2014)

Parameters: $(\eta_t)_t > 0$, (++).

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

For each round $t = 1,2, \dots, T$

- For all i, let

$$p_{t,i} = \left(1 - \sum_j \varepsilon_{t,j}\right) \frac{w_{t,i}}{\sum_j w_{t,j}} + \varepsilon_{t,i}.$$

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

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

- For all i, update weight as

$$w_{t+1,i} = \exp\left(-\eta_t \hat{L}_{t,i}\right)$$

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

Theorem (SS, 2014):
The regret of **Exp3++** satisfies

$$\hat{R}_T \leq 4\sqrt{TK \log K}$$

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

Theorem (SS, 2014):
The regret of **Exp3++** satisfies

$$\hat{R}_T \leq 4\sqrt{TK \log K}$$

Proof idea: the $\varepsilon_{t,i}$'s are small enough to not change the standard **Exp3** analysis:

$$\varepsilon_{t,i} = O\left(\sqrt{\log K \,/KT}\right)$$

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

Theorem (SS, 2014):
The regret of **Exp3++** satisfies

$$\hat{R}_T = O\left(\tilde{C}(\nu)\log^3 T + C'(\nu)\right)$$

in the stochastic case

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

Proof ideas:

- Let $\Delta_i = \mu_i - \mu^*$

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

Proof ideas:

- Let $\Delta_i = \mu_i - \mu^*$
- **Wishful thinking:** if we had full information, then

$$p_{t,i} \approx \frac{e^{-t\eta_t\Delta_i}}{\sum_j e^{-t\eta_t\Delta_j}} \leq e^{-t\eta_t\Delta_i}$$

holds for all suboptimal arms i

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

Proof ideas:

- Let $\Delta_i = \mu_i - \mu^*$
- **Wishful thinking:** if we had full information, then

$$p_{t,i} \approx \frac{e^{-t\eta_t\Delta_i}}{\sum_j e^{-t\eta_t\Delta_j}} \leq e^{-t\eta_t\Delta_i}$$

  holds for all suboptimal arms i
- Thus, the expected number of suboptimal draws is

$$\sum_{t=1}^{T} p_{t,i} \leq \sum_{t=1}^{T} e^{-t\eta_t\Delta_i} = O\left(\frac{K}{\Delta_i^2}\right)$$

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

But we don't have full info :(

Proof ideas:

- Let $\Delta_i = \mu_i - \mu^*$
- **Wishful thinking:** if we had full information, then

$$p_{t,i} \approx \frac{e^{-t\eta_t\Delta_i}}{\sum_j e^{-t\eta_t\Delta_j}} \leq e^{-t\eta_t\Delta_i}$$

  holds for all suboptimal arms i
- Thus, the expected number of suboptimal draws is

$$\sum_{t=1}^{T} p_{t,i} \leq \sum_{t=1}^{T} e^{-t\eta_t\Delta_i} = O\left(\frac{K}{\Delta_i^2}\right)$$

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

- **Idea:** ensure that the estimated gap is "reasonable":

$$t\widehat{\Delta}_{t,i} \stackrel{\text{def}}{=} \widehat{L}_{t,i} - \widehat{L}_t^* \geq t\Delta_i - o(t)$$

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

ensured by the exploration parameters $\varepsilon_{t,i}$!!!

- **Idea:** ensure that the estimated gap is "reasonable":

$$t\widehat{\Delta}_{t,i} \stackrel{\text{def}}{=} \widehat{L}_{t,i} - \widehat{L}_t^* \geq t\Delta_i - o(t)$$

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

ensured by the exploration parameters $\varepsilon_{t,i}$!!!

- **Idea:** ensure that the estimated gap is "reasonable":

$$t\widehat{\Delta}_{t,i} \stackrel{\text{def}}{=} \widehat{L}_{t,i} - \widehat{L}_t^* \geq t\Delta_i - o(t)$$

- For large enough t ($t \geq t^*$), we have $t\widehat{\Delta}_{t,i} \geq t\Delta_i/2$

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

ensured by the exploration parameters $\varepsilon_{t,i}$!!!

- **Idea:** ensure that the estimated gap is "reasonable":

$$t\widehat{\Delta}_{t,i} \stackrel{\text{def}}{=} \widehat{L}_{t,i} - \widehat{L}_t^* \geq t\Delta_i - o(t)$$

- For large enough t ($t \geq t^*$), we have $t\widehat{\Delta}_{t,i} \geq t\Delta_i/2$
- This gives

$$p_{t,i} = \frac{e^{-t\eta_t\widehat{\Delta}_{t,i}}}{\sum_j e^{-t\eta_t\widehat{\Delta}_{t,j}}} \leq e^{-t\eta_t\widehat{\Delta}_{t,i}} \leq e^{-t\eta_t\Delta_i/2}$$

for all suboptimal arms i

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

ensured by the exploration parameters $\varepsilon_{t,i}$!!!

- **Idea:** ensure that the estimated gap is "reasonable":

$$t\widehat{\Delta}_{t,i} \stackrel{\text{def}}{=} \widehat{L}_{t,i} - \widehat{L}_t^* \geq t\Delta_i - o(t)$$

- For large enough t ($t \geq t^*$), we have $t\widehat{\Delta}_{t,i} \geq t\Delta_i/2$
- This gives

$$p_{t,i} = \frac{e^{-t\eta_t\widehat{\Delta}_{t,i}}}{\sum_j e^{-t\eta_t\widehat{\Delta}_{t,j}}} \leq e^{-t\eta_t\widehat{\Delta}_{t,i}} \leq e^{-t\eta_t\Delta_i/2}$$

  for all suboptimal arms i

- Thus,

$$\sum_{t=1}^{T} p_{t,i} \leq t^* + \sum_{t=t^*}^{T} e^{-t\eta_t\Delta_i/2} = t^* + O\left(\frac{K}{\Delta_i^2}\right)$$

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

ensured by the exploration parameters $\varepsilon_{t,i}$!!!

- **Idea:** ensure that the estimated gap is "reasonable":

$$t\widehat{\Delta}_{t,i} \stackrel{\text{def}}{=} \widehat{L}_{t,i} - \widehat{L}_t^* \geq t\Delta_i - o(t)$$

- For large enough t ($t \geq t^*$), we have $t\widehat{\Delta}_{t,i} \geq t\Delta_i/2$
- This gives

$$p_{t,i} = \frac{e^{-t\eta_t\widehat{\Delta}_{t,i}}}{\sum_j e^{-t\eta_t\widehat{\Delta}_{t,j}}} \leq e^{-t\eta_t\widehat{\Delta}_{t,i}} \leq e^{-t\eta_t\Delta_i/2}$$

  for all suboptimal arms i

The rest is grinding out the asymptotics…

- Thus,

$$\sum_{t=1}^{T} p_{t,i} \leq t^* + \sum_{t=t^*}^{T} e^{-t\eta_t\Delta_i/2} = t^* + O\left(\frac{K}{\Delta_i^2}\right)$$

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

Bottom line:
"if there is a linear gap between $L_{t,i}$ and L_t^*, this should be exposed in the estimated gap $t\widehat{\Delta}_{t,i}$"

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

Bottom line:
"if there is a linear gap between $L_{t,i}$ and L_t^*, this should be exposed in the estimated gap $t\widehat{\Delta}_{t,i}$"

Corollaries: strong bounds whenever there is such a gap:

- "contaminated stochastic"
- "adversarial with a gap"

EXP3++ ANALYSIS (HEAVILY PARAPHRASED)

Bottom line:
"if there is a linear gap between $L_{t,i}$ and L_t^*, this should be exposed in the estimated gap $t\widehat{\Delta}_{t,i}$"

That's the exact opposite of what we need for 1st order bounds!

- "adversarial with a gap"

OPEN QUESTIONS

Is there a way to exploit gaps that are growing slower than linear?

Is there a way to improve asymptotics? (In SS'14, t^* is horribly big!)

OPEN QUESTIONS

Is there a way to exploit gaps that are growing slower than linear?

Is there a way to improve asymptotics? (In SS'14, t^* is horribly big!)

So far, all positive results hold only for oblivious adversaries—is it possible to extend these to adaptive ones?

OPEN QUESTIONS

Is there a way to exploit gaps that are growing slower than linear?

Is there a way to improve asymptotics? (In SS'14, t^* is horribly big!)

So far, all positive results hold only for oblivious adversaries—is it possible to extend these to adaptive ones?

See Peter's talk tomorrow!

BEYOND MINIMAX #3: PRIOR-DEPENDENT BOUNDS

PRIOR-DEPENDENT BOUNDS FOR FULL INFO

Theorem

(Luo and Schapire, 2015, Koolen and Van Erven, 2015, Orabona and Pal, 2016)

There exist algorithms guaranteeing

$$\hat{R}_T(\rho) = O\left(\sqrt{T\left(1 + \mathrm{RE}(\rho|\pi)\right)}\right)$$

for any fixed prior $\pi \in \Delta_K$ and any comparator $\rho \in \Delta_K$

Theorem

(Even-Dar et al., 2007, Sani et al., 2014)

There exist algorithms guaranteeing

$$\hat{R}_T(i) = const$$

for any fixed i, while also guaranteeing

$$\hat{R}_T = \tilde{O}\left(\sqrt{T}\right)$$

PRIOR-DEPENDENT BOUNDS FOR FULL INFO

Theorem

(Luo and Schapire, 2015, Koolen and Van Erven, 2015, Orabona and Pal, 2016)

uaranteeing

$$\left(\mathrm{KL}(\rho|\pi)\right)$$

for an y comparator $\rho \in \Delta_K$

Anything similar possible for bandits??

Theorem

(Even-Dar et al., 2007, Sani et al., 2014)

There exist algorithms guaranteeing

$$\hat{R}_T(i) = const$$

for any fixed i, while also guaranteeing

$$\hat{R}_T = \tilde{O}\left(\sqrt{T}\right)$$

PRIOR-DEPENDENT BOUNDS FOR FULL INFO

Theorem

(Luo and Schapire, 2015, Koolen and Van Erven, 2015, Orabona and Pal, 2016)

uaranteeing

$\mathrm{KL}(\rho|\pi))\Big)$

for an y comparator $\rho \in \Delta_K$

Anything similar possible for bandits??

NO* :(

*not quite

(Even-Dar e

There exist a ng

$\hat{R}$

for any fixed i, while also guaranteeing

$\hat{R}_T = \tilde{O}(\sqrt{T})$

PRIOR-DEPENDENT BOUNDS FOR BANDITS

Theorem (Lattimore, 2015) paraphrased
The regrets $\hat{R}_T(i)$ need to satisfy

$$\hat{R}_T(i) \geq \min\left\{T, \sum_{j \neq i} \frac{T}{\hat{R}_T(j)}\right\}.$$

In particular,

- $\hat{R}_T(i) = const$ implies $\hat{R}_T(j) = \Omega(T)$
- Fixing a prior π and getting a bound $\hat{R}_T(\rho) = \tilde{O}\left(\sqrt{T \sum_j (\rho_j / \pi_j)}\right)$ is not possible

PRIOR-DEPENDENT BOUNDS: "POSITIVE" RESULTS

Lattimore (2015):

- For any regret bound satisfying the condition, there exists an algorithm achieving it in the stochastic setting
- In particular, $\sum_j \frac{\rho_j}{\pi_j}\sqrt{T}$ is achievable (see also Rosin, 2011)

Neu (2016, made up on the flight here):

- For non-stochastic bandits, there is an algorithm with

$$\hat{R}_T(i) = \tilde{O}\left(\sqrt{\frac{KT\ \mathrm{softmax}(\pi)}{\pi_i}}\right)$$

BEYOND MINIMAX: CONCLUSIONS

CONCLUSIONS

Higher order bounds

- First-order bounds are possible like in full info
- Second order bounds: much weaker than full info

Best-of-both-world bounds

- Possible and strong against oblivious adversaries
- Only weak guarantees for adaptive adversaries

Prior dependent bounds

- Nothing fancy is possible


$1,015,150.50
THANKS!!!