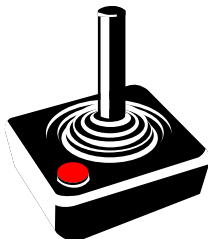


# Maximin Action Identification



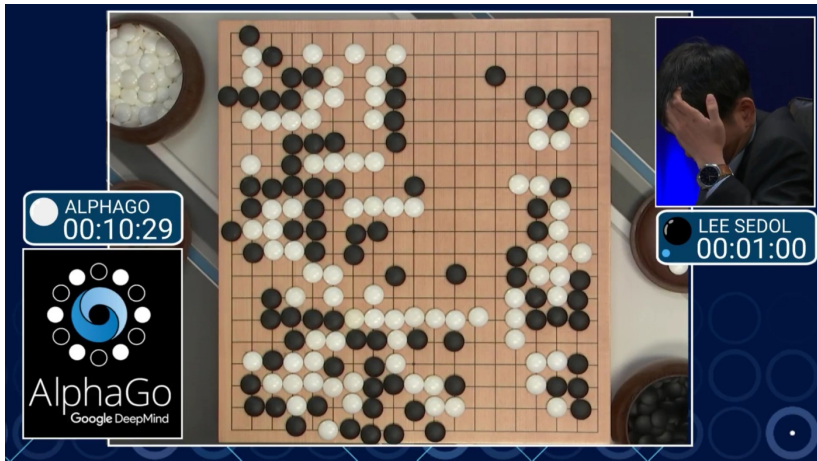
Wouter M. Koolen



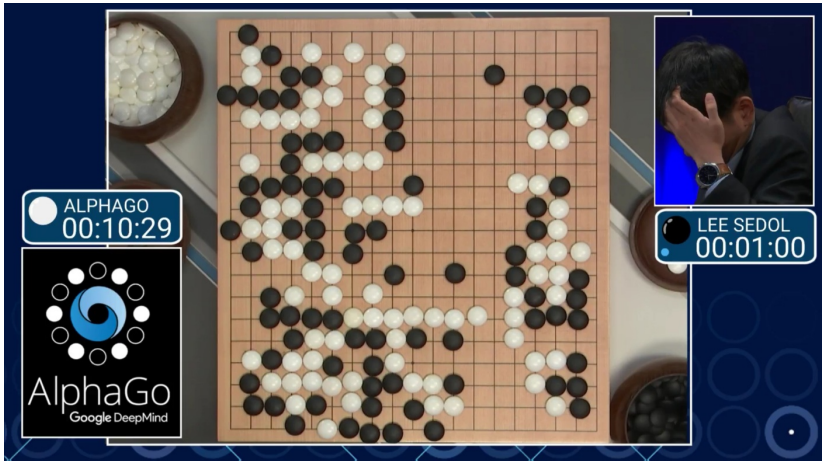
Theoretical Foundations for Learning from Easy Data  
Leiden, Friday 11<sup>th</sup> November, 2016

Joint work with Aurélien Garivier and Emilie Kaufman

But can we make it work in Theory?

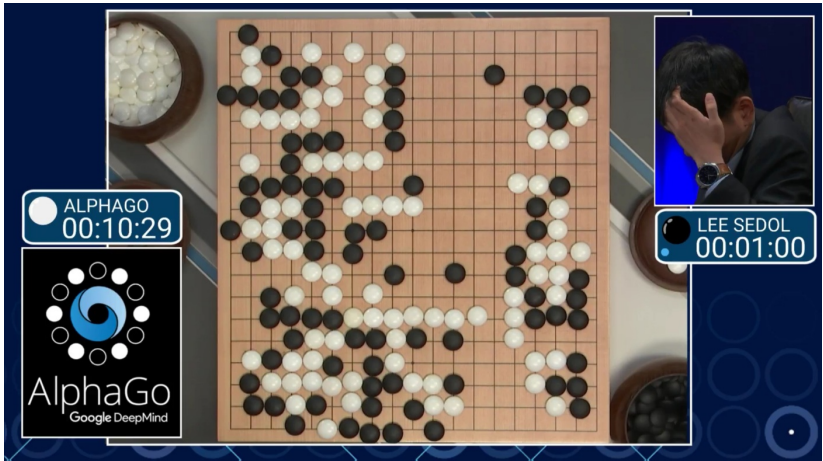


But can we make it work in Theory?



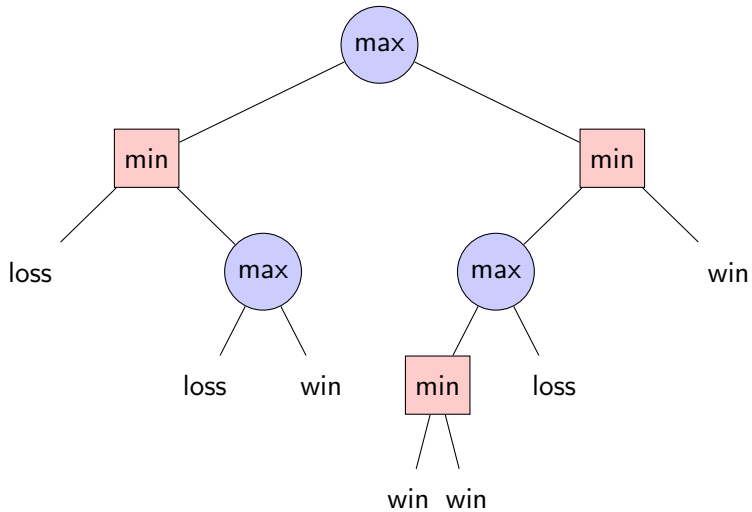
Deep Neural Networks + Monte Carlo Tree Search

But can we make it work in Theory?

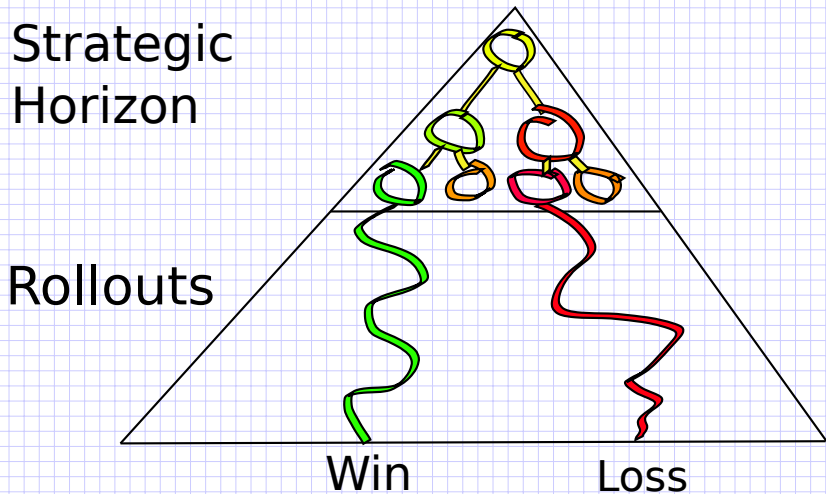


Deep Neural Networks + **Monte Carlo Tree Search**

# Game Tree



# Monte Carlo Tree Search (MCTS)



## Main Question

What is the sample complexity of MCTS?

## Main Question

What is the sample complexity of MCTS?

Which game trees are **easy**?



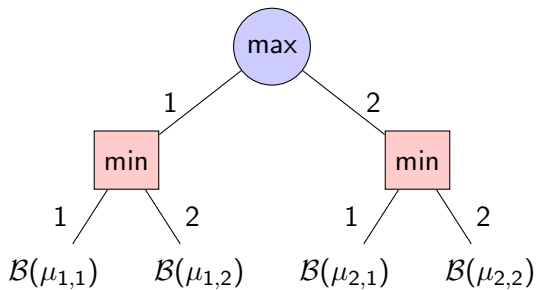
## Main Question

What is the sample complexity of MCTS?

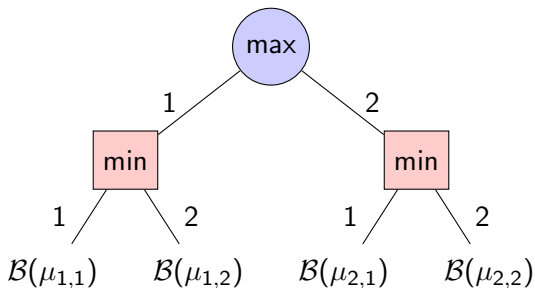
Which game trees are **easy**?

End-to-end **economy**.

## Simplest Model



## Simplest Model



Goal: find **maximin action**

$$i^* := \arg \max_i \min_j \mu_{i,j}$$

# Protocol

Strategy for maximin sample identification

- ▶ Sampling rule  $(I_t, J_t)$  (observation  $X_t \sim \mathcal{B}(\mu_{I_t, J_t})$ )
- ▶ Stopping rule  $\tau$
- ▶ Recommendation rule  $\hat{i}$ .

# Protocol

Strategy for maximin sample identification

- ▶ Sampling rule  $(I_t, J_t)$  (observation  $X_t \sim \mathcal{B}(\mu_{I_t, J_t})$ )
- ▶ Stopping rule  $\tau$
- ▶ Recommendation rule  $\hat{i}$ .

Two criteria

Time

$$\mathbb{E}_{\mu}[\tau]$$

Quality

$$\mathbb{P}_{\mu}(\hat{i} \neq i^*)$$

# Fixed Confidence Setting

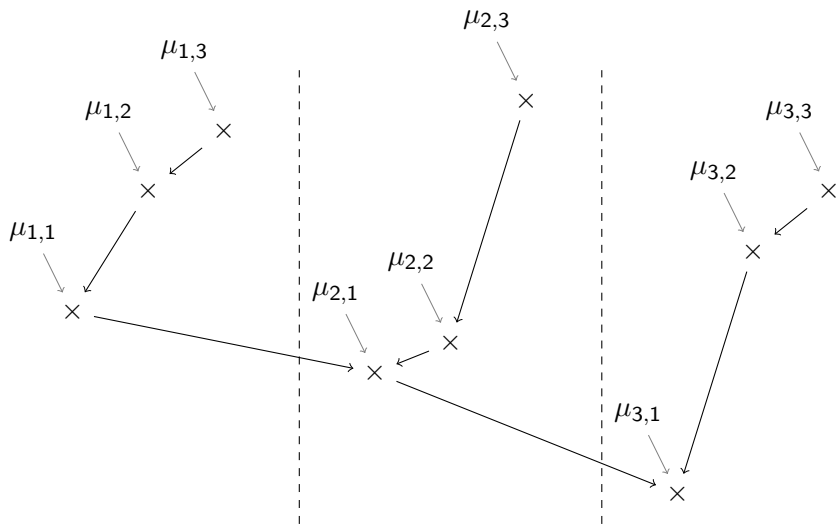
## Definition

A strategy  $(P_t, \tau, \hat{i})$  is  $\delta$ -**PAC** if for all\*  $\boldsymbol{\mu} = \{\mu_{i,j}\}$

$$\mathbb{P}_{\boldsymbol{\mu}}(\hat{i} \neq i^*) \leq \delta.$$

Goal: minimize **sample complexity**  $\mathbb{E}_{\boldsymbol{\mu}} \tau$

# Normal Form



## Squelch our Inner Reductionist

Reduction to Best Arm Identification (BAI):

- ▶ Apply BAI **hierarchically**



## Squelch our Inner Reductionist

Reduction to Best Arm Identification (BAI):

- ▶ Apply BAI **hierarchically**
- ▶ Use BAI to find **worst leaf** (for 2 max moves)

## Squelch our Inner Reductionist

Reduction to Best Arm Identification (BAI):

- ▶ Apply BAI **hierarchically**
- ▶ Use BAI to find **worst leaf** (for 2 max moves)

$$\frac{\mathbb{E}_{\mu}[\tau]}{\ln(1/\delta)} \asymp \frac{2}{(\mu_{1,1} - \mu_{2,1})^2} + \frac{1}{(\mu_{1,2} - \mu_{2,1})^2} + \frac{1}{(\mu_{2,2} - \mu_{2,1})^2}.$$

## Squelch our Inner Reductionist

Reduction to Best Arm Identification (BAI):

- ▶ Apply BAI **hierarchically**
- ▶ Use BAI to find **worst leaf** (for 2 max moves)

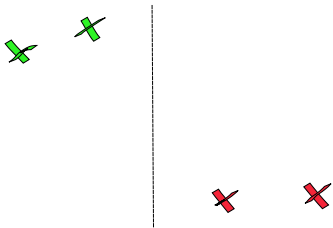
$$\frac{\mathbb{E}_{\mu}[\tau]}{\ln(1/\delta)} \asymp \frac{2}{(\mu_{1,1} - \mu_{2,1})^2} + \frac{1}{(\mu_{1,2} - \mu_{2,1})^2} + \frac{1}{(\mu_{2,2} - \mu_{2,1})^2}.$$

# Squelch our Inner Reductionist

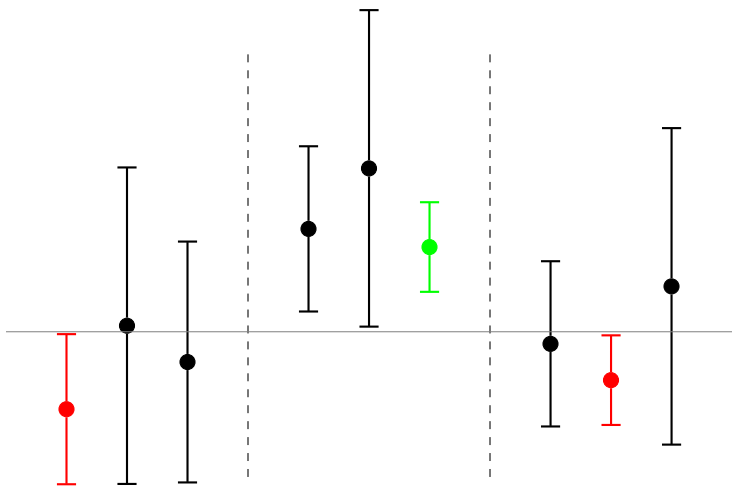
Reduction to Best Arm Identification (BAI):

- ▶ Apply BAI **hierarchically**
- ▶ Use BAI to find **worst leaf** (for 2 max moves)

$$\frac{\mathbb{E}_{\mu}[\tau]}{\ln(1/\delta)} \asymp \frac{2}{(\mu_{1,1} - \mu_{2,1})^2} + \frac{1}{(\mu_{1,2} - \mu_{2,1})^2} + \frac{1}{(\mu_{2,2} - \mu_{2,1})^2}.$$



# M-LUCB Algorithm



# M-LUCB

- ▶ Maintain confidence interval  $[L_{i,j}, U_{i,j}]$  for each leaf.
- ▶ Pick **representative** for each action

$$j_i = \arg \min_j L_{i,j}$$

- ▶ BAI step. With empirical maximin

$$\hat{i} = \arg \max_i \min_j \hat{\mu}_{i,j}$$

and “contender”

$$\tilde{i} = \arg \max_{i \neq \hat{i}} U_{i,j_i}$$

draw both

$$X := (\hat{i}, j_{\hat{i}}) \quad \text{and} \quad C := (\tilde{i}, j_{\tilde{i}}).$$

- ▶ Stop when

$$L_X > U_C$$

## M-LUCB Sample Complexity

Let

$$L_{i,j} = \hat{\mu}_{i,j} - \sqrt{\frac{\beta(t, \delta)}{2N_{i,j}}} \quad \text{and} \quad U_{i,j} = \hat{\mu}_{i,j} + \sqrt{\frac{\beta(t, \delta)}{2N_{i,j}}}$$

### Theorem

Let  $\alpha > 1$ . There is  $C > 0$  such that for

$$\beta(t, \delta) = \ln(Ct^{1+\alpha}/\delta)$$

M-LUCB is  $\delta$ -PAC and

$$\limsup_{\delta \rightarrow 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau]}{\ln(1/\delta)} \leq 8(1 + \alpha)H^*(\boldsymbol{\mu})$$

where

$$H^*(\boldsymbol{\mu}) = \sum_j \frac{1}{(\mu_{1,j} - \mu_{2,1})^2} + \sum_{i>1} \sum_j \frac{1}{(\mu_{1,1} - \mu_{i,1})^2 \vee (\mu_{i,j} - \mu_{i,1})^2}$$

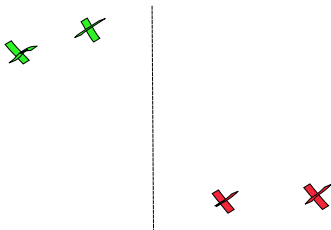
## M-LUCB Good?

BAI for  $2 \times 2$  game gave

$$\frac{\mathbb{E}_{\mu}[\tau]}{\ln(1/\delta)} \asymp \frac{2}{(\mu_{1,1} - \mu_{2,1})^2} + \frac{1}{(\mu_{1,2} - \mu_{2,1})^2} + \frac{1}{(\mu_{2,2} - \mu_{2,1})^2}$$

whereas M-LUCB gives

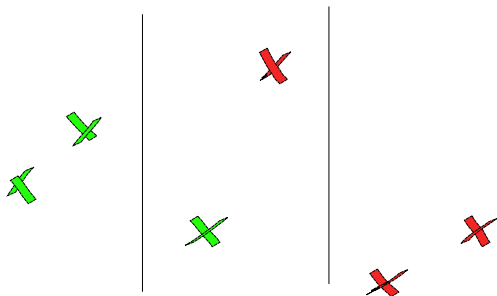
$$\frac{\mathbb{E}_{\mu}[\tau]}{\ln(1/\delta)} \asymp \frac{2}{(\mu_{1,1} - \mu_{2,1})^2} + \frac{1}{(\mu_{1,2} - \mu_{2,1})^2} + \frac{1}{(\mu_{1,1} - \mu_{2,1})^2 \vee (\mu_{2,2} - \mu_{2,1})^2}$$





# M-Racing Algorithm

- ▶ Idea: successively eliminate leaves.
- ▶ Each epoch we pull all remaining leaves.
- ▶ Eliminate **high leaf**  $(i, j)$  if  $\exists j'$  such that  $\hat{\mu}_{i,j} \gg \hat{\mu}_{i,j'}$ .
- ▶ Eliminate **low action**  $i$  if  $\exists i'$  such that  $\min_j \hat{\mu}_{i',j} \ll \min_j \hat{\mu}_{i,j}$ .



## M-Racing Sample Complexity

Divergence function:

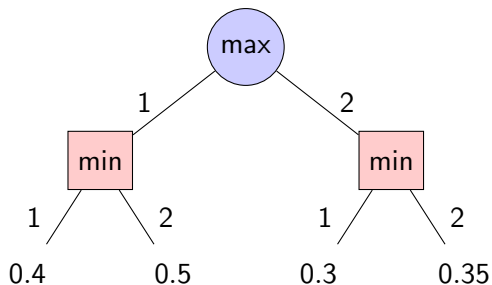
$$I(\mu, \nu) = \left[ \text{KL} \left( \mu, \frac{\mu + \nu}{2} \right) + \text{KL} \left( \nu, \frac{\mu + \nu}{2} \right) \right] \mathbf{1}_{\mu \geq \nu}$$

Pinsker:  $I(\mu, \nu) \geq (\mu - \nu)^2$

M-Racing guarantees

$$\frac{\mathbb{E}_{\mu}[\tau]}{\ln(1/\delta)} \asymp \sum_j \frac{1}{I(\mu_{1,1}, \mu_{2,1}) \vee I(\mu_{1,j}, \mu_{1,1})} + \sum_{i>1,j} \frac{1}{I(\mu_{i,1}, \mu_{1,1}) \vee I(\mu_{i,j}, \mu_{i,1})}$$

## Experiments



	$\tau_{1,1}$	$\tau_{1,2}$	$\tau_{2,1}$	$\tau_{2,2}$	sum
M-LUCB	1762	198	1761	462	4183
M-KL-LUCB	762	92	733	237	1824
M-Chernoff	<b>315</b>	<b>59</b>	<b>291</b>	<b>136</b>	<b>801</b>
M-Racing	324	152	301	298	1075
KL-LUCB	351	64	3074	2768	6257

# Lower Bound

## Theorem

Any  $\delta$ -PAC algorithm satisfies

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau_{\delta}] \geq T^*(\boldsymbol{\mu}) d(\delta, 1 - \delta),$$

where

$$T^*(\boldsymbol{\mu})^{-1} := \sup_{w \in \Delta} \inf_{\boldsymbol{\mu}': \mu'_{1,1} \wedge \mu'_{1,2} < \mu'_{2,1} \wedge \mu'_{2,2}} \left( \sum_{i,j} w_{i,j} d(\mu_{i,j}, \mu'_{i,j}) \right)$$

For BAI such lower bounds lead to optimal algorithms [Garivier and Kaufmann, 2016].

For Maximin Action Identification we cannot prove that you have to pull arm (2, 2) linearly often when  $\mu_{2,2} > \mu_{1,2}$ .

# Conclusion



## Maximin Action Identification

- ▶ Interesting and challenging problem
- ▶ First set of algorithms
- ▶ Sample complexity guarantees

## Future work

- ▶ Characterization of optimal strategies
- ▶ Depth  $> 2$
- ▶ Fixed budget setting

## Open Career Path

Given  $f : \mathbb{R}^d \rightarrow \mathcal{I}$ . What is the sample complexity of estimating

$$f(\mathbb{E} X_1, \dots, \mathbb{E} X_d)?$$

This talk: maximin action

$$f(\{\mu_{i,j}\}) = \arg \max_i \min_j \mu_{i,j}$$

with finite  $\mathcal{I}$ .

Interesting variation: best mixed strategy for matrix game

$$f(\{\mu_{i,j}\}) = \arg \max_{p \in \Delta} \min_j \sum_i p_i \mu_{i,j}$$

with simplex  $\mathcal{I}$ .