Extreme Value Monte Carlo Tree Search

Masataro Asai[1] and Stephen Wissow[2]

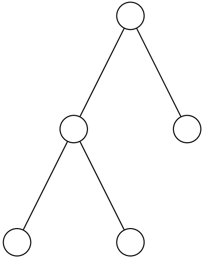
[1]MIT-IBM Watson AI Lab, USA, [2]University of New Hampshire, USA

This work was supported through DTIC contract FA8075-18-D-0008, Task Order FA807520F0060, Task 4 - Autonomous Defensive Cyber Operations (DCO) Research & Development (R&D).

Monte Carlo Tree Search (MCTS/UCT)

Auer, Cesa-Bianchi, and Fischer (2002, UCB1), Kocsis and Szepesvári (2006, UCT), Schulte and Keller (2014, THTS)

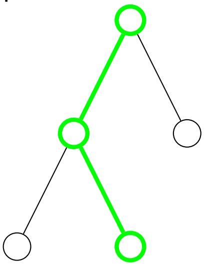
partial tree, mid-search:



Monte Carlo Tree Search (MCTS/UCT)

Auer, Cesa-Bianchi, and Fischer (2002, UCB1), Kocsis and Szepesvári (2006, UCT), Schulte and Keller (2014, THTS)

partial tree, mid-search:

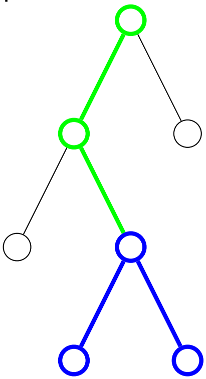


1. **select** a leaf node by
   $\mathrm{LCB1}_i = \hat{\mu}_i - c\sqrt{\frac{2\log T}{t_i}}$
   $\hat{\mu}_i$: mean of child i
   t_i : visit count of i
   T : parent visit count

Monte Carlo Tree Search (MCTS/UCT)

Auer, Cesa-Bianchi, and Fischer (2002, UCB1), Kocsis and Szepesvári (2006, UCT), Schulte and Keller (2014, THTS)

partial tree, mid-search:

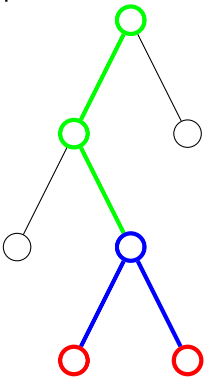


1. **select** a leaf node by
   $\text{LCB1}_i = \hat{\mu}_i - c\sqrt{\frac{2\log T}{t_i}}$
   $\hat{\mu}_i$: mean of child i
   t_i : visit count of i
   T : parent visit count
2. **expand** a leaf node

Monte Carlo Tree Search (MCTS/UCT)

Auer, Cesa-Bianchi, and Fischer (2002, UCB1), Kocsis and Szepesvári (2006, UCT), Schulte and Keller (2014, THTS)

partial tree, mid-search:

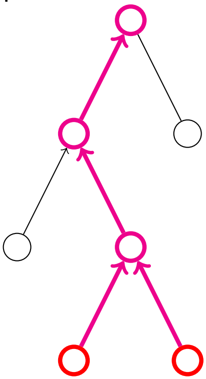


1. **select** a leaf node by
   $\text{LCB1}_i = \hat{\mu}_i - c\sqrt{\frac{2\log T}{t_i}}$
   $\hat{\mu}_i$: mean of child i
   t_i : visit count of i
   T : parent visit count
2. **expand** a leaf node
3. **evaluate** heuristics of children

Monte Carlo Tree Search (MCTS/UCT)

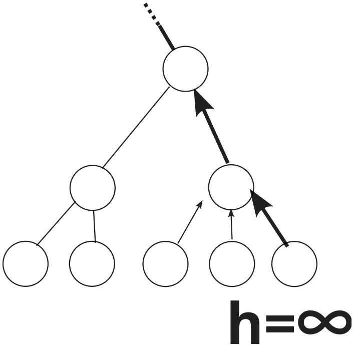
Auer, Cesa-Bianchi, and Fischer (2002, UCB1), Kocsis and Szepesvári (2006, UCT), Schulte and Keller (2014, THTS)

partial tree, mid-search:

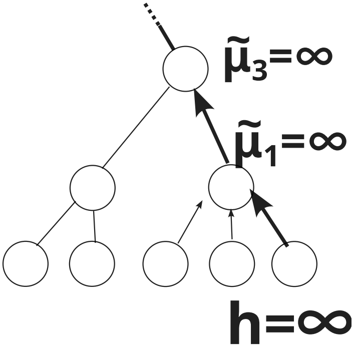


1. **select** a leaf node by
   $\text{LCB1}_i = \hat{\mu}_i - c\sqrt{\frac{2\log T}{t_i}}$
   $\hat{\mu}_i$: mean of child i
   t_i : visit count of i
   T : parent visit count
2. **expand** a leaf node
3. **evaluate** heuristics of children
4. **backup** information to ancestors
   $t = \sum_i t_i$: Sum of children
   $\hat{\mu} = \frac{\sum_i t_i \hat{\mu}_i}{\sum_i t_i}$: Weighted avg

Average is Weird in Planning: What happens at a dead-end?

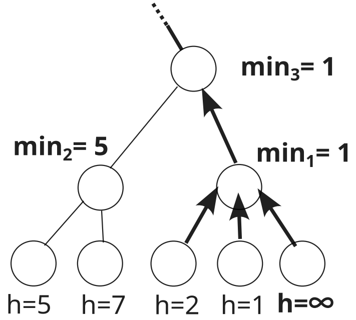


Average is Weird in Planning: What happens at a dead-end?



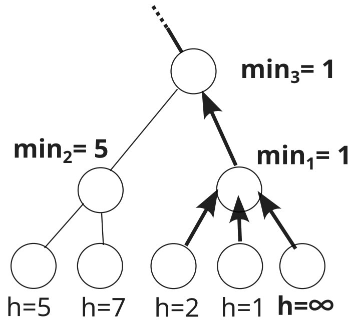
- **A single $h = \infty \rightarrow$ all $\hat{\mu} = \infty$**

Average is Weird in Planning: What happens at a dead-end?



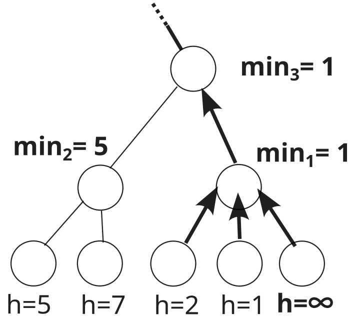
- **A single $h = \infty \rightarrow$ all $\hat{\mu} = \infty$**
- $\min$ has no such problem

Average is Weird in Planning: What happens at a dead-end?



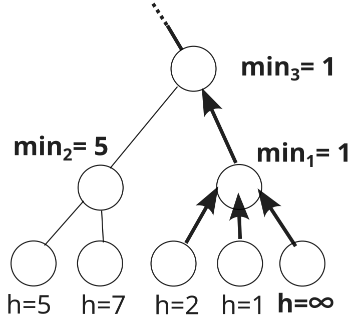
- **A single $h = \infty \to$ all $\hat{\mu} = \infty$**
- $\min$ has no such problem
- backup min h, select min h = GBFS
  backup avg h, select UCB1 = GUCT
  backup min h, select UCB1 = GUCT*
  (Schulte and Keller 2014, THTS)
  However...

Average is Weird in Planning: What happens at a dead-end?



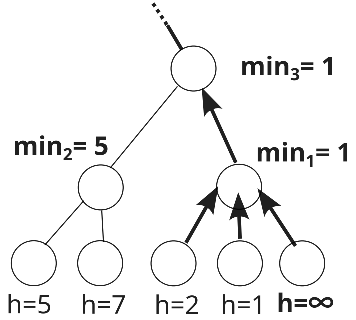
- **A single $h = \infty \rightarrow$ all $\hat{\mu} = \infty$**
- $\min$ has no such problem
- backup min h, select min h = GBFS
  backup avg h, select UCB1 = GUCT
  backup min h, select UCB1 = GUCT*
  (Schulte and Keller 2014, THTS)
  However...
- $\min$ **lacks statistical interpretation**

Average is Weird in Planning: What happens at a dead-end?



- **A single $h = \infty \rightarrow$ all $\hat{\mu} = \infty$**
- $\min$ has no such problem
- backup min h, select min h = GBFS
  backup avg h, select UCB1 = GUCT
  backup min h, select UCB1 = GUCT*
  (Schulte and Keller 2014, THTS)
  However...
- **$\min$ lacks statistical interpretation**
- GUCT/* also remove ∞ from the tree

Average is Weird in Planning: What happens at a dead-end?



- **A single $h = \infty \to$ all $\hat{\mu} = \infty$**
- $\min$ has no such problem
- backup min h, select min h = GBFS
  backup avg h, select UCB1 = GUCT
  backup min h, select UCB1 = GUCT*
  (Schulte and Keller 2014, THTS)
  However...
- **$\min$ lacks statistical interpretation**
- GUCT/* also remove ∞ from the tree
- **Lacks statistical interpretation**

The key slide™

The average is weird, but how to use the minimum with statistical rigor?

The key slide™

The average is weird, but how to use the minimum with statistical rigor?

The statistical theory of the average:

is **Central Limit Theorem (CLT).**

The key slide™

The average is weird, but how to use the minimum with statistical rigor?

The statistical theory of the average:
is **Central Limit Theorem (CLT).**

The statistical theory of the minimum (or maximum)?

The key slide™

The average is weird, but how to use the minimum with statistical rigor?

The statistical theory of the average:
is **Central Limit Theorem (CLT).**

The statistical theory of the minimum (or maximum)?
It's **Extreme Value Theory (EVT)!**

Extreme Value Theory (EVT)

Safety-critical applications: e.g. **Maximum water level in rivers**

Extreme Value Theory (EVT)

Safety-critical applications: e.g. **Maximum water level in rivers**

- **Monthly average water level**

Extreme Value Theory (EVT)

Safety-critical applications: e.g. **Maximum water level in rivers**

- **Monthly average water level**
  $\rightarrow$ Gaussian distribution

Extreme Value Theory (EVT)

Safety-critical applications: e.g. **Maximum water level in rivers**

- **Monthly average water level**
  $\rightarrow$ Gaussian distribution
- **Exceedance over the safety limit**

Extreme Value Theory (EVT)

Safety-critical applications: e.g. **Maximum water level in rivers**

- **Monthly average water level**
  $\rightarrow$ Gaussian distribution
- **Exceedance over the safety limit**
  $\rightarrow$ **Generalized Pareto (GP) distribution**

Extreme Value Theory (EVT)

Safety-critical applications: e.g. **Maximum water level in rivers**

- **Monthly average water level**
  → Gaussian distribution
- **Exceedance over the safety limit**
  → **Generalized Pareto (GP) distribution**

$$\mathrm{GP}(x \mid \theta, \sigma, \xi) = \text{Unimportant, complicated math.} \quad (x > \theta)^{*}$$ **for threshold θ**

Extreme Value Theory (EVT)

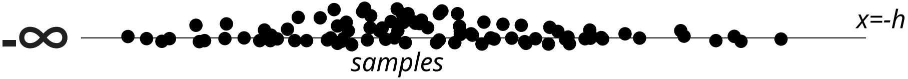
Safety-critical applications: e.g. **Maximum water level in rivers**

- **Monthly average water level**
  $\rightarrow$ Gaussian distribution
- **Exceedance over the safety limit**
  $\rightarrow$ **Generalized Pareto (GP) distribution**

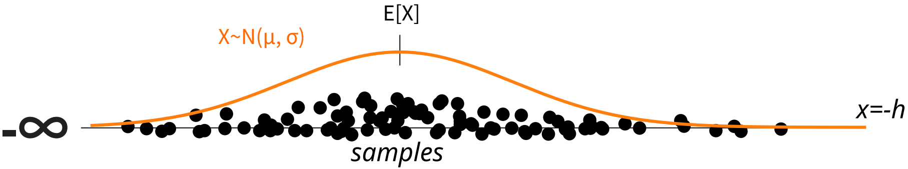
$$\mathrm{GP}(\mathsf{x} \mid \theta, \sigma, \xi) = \text{Unimportant, complicated math.} \quad (\mathsf{x} > \theta)^{*}$$ **for threshold θ**

***justifies removing dead-ends:** $x = -h = -\infty$ (minimize h; maximize x)

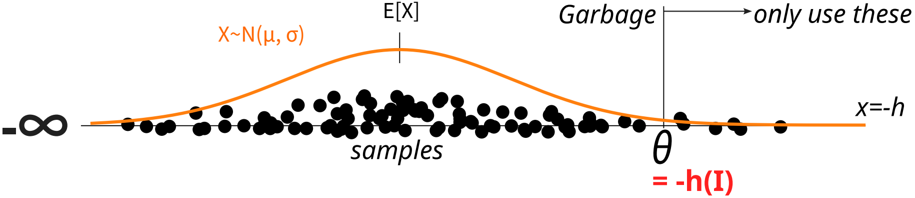
Backup = Fitting a distribution



Backup = Fitting a distribution

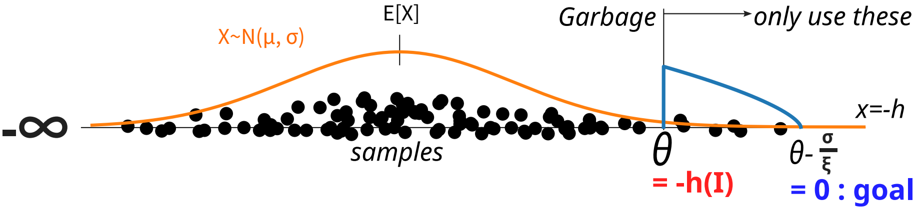


Backup = Fitting a distribution



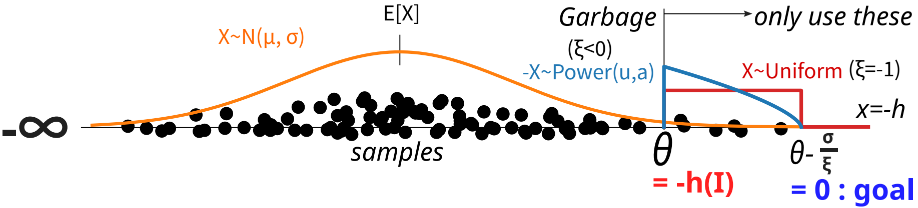
- **initial state** I

Backup = Fitting a distribution



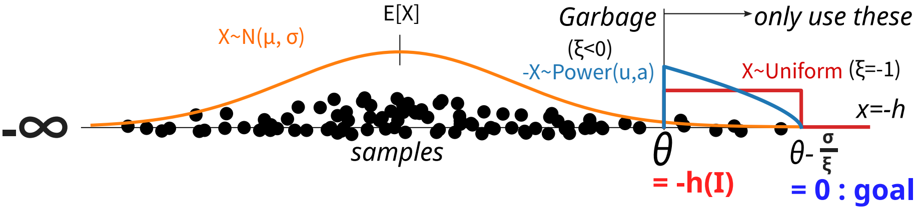
- **initial state I**
- We fit **exceedance over $\theta = -h(I)$ to $\mathrm{GP}(\theta, \sigma, \xi)$**

Backup = Fitting a distribution



- **initial state I**
- We fit **exceedance over $\theta = -h(I)$ to $\mathrm{GP}(\theta, \sigma, \xi)$**
- We use special cases of GP: Power $(\xi < 0)$ and Uniform $(\xi = -1)$

Backup = Fitting a distribution



- **initial state** I
- We fit **exceedance over** $\theta = -h(I)$ **to** $\mathrm{GP}(\theta, \sigma, \xi)$
- We use special cases of GP: Power $(\xi < 0)$ and Uniform $(\xi = -1)$
- We define two Bandits:

$$
\begin{aligned}
\text{LCB1-Uniform}_i &= \frac{\hat{u}_i + \hat{l}_i}{2} - (\hat{u}_i - \hat{l}_i)\sqrt{6 t_i \log T} \\
\text{LCB1-Power}_i &= \frac{\hat{u}_i \hat{a}_i}{\hat{a}_i + 1} - \hat{u}_i \sqrt{6 t_i \log T}
\end{aligned}
$$

Results

Num. solved on 24 IPC domains w/ 10^4 evaluations

$h =$	h^{FF}	h^{add}	$h^{\max}$	h^{GC}	h^{FF}+PO	h^{FF}+DE	h^{FF}+DE+PO
GBFS	538	518	224	354	-	489	-
Softmin-Type(h)	576	542.6	297.2	357.6	-	578	-
GUCT **Uses UCB1**	412	397.8	228.4	285.2	454	389.2	439.4
-Normal **Uses UCB1-Normal**	283.4	265	212	233.4	372.4	289	381.6
*-Normal **+ backprop min**	318.8	300	215.2	246.2	378.05	304.4	386.7
-Normal2	581.8	535.8	316.6	379	621	518	578
*-Normal2	567.2	533.8	263	341	618	511.4	567.8
-Power	**596**	541.8	**450.6**	463.2	**623.4**	413.6	**583**
-Uniform	**594.8**	**543.8**	**450.6**	**463.8**	**626.4**	416.4	**583**

FD/C++ implementation is on the way and showing promising results

Find our full paper: https://arxiv.org/abs/2405.18248

References I

Auer, P.; Cesa-Bianchi, N.; and Fischer, P. 2002. Finite-Time Analysis of the Multiarmed Bandit Problem. *Machine Learning*, 47(2-3): 235–256.

Kocsis, L.; and Szepesvári, C. 2006. Bandit Based Monte-Carlo Planning. In *Proc. of ECML*, 282–293. Springer.

Schulte, T.; and Keller, T. 2014. Balancing Exploration and Exploitation in Classical Planning. In *Proc. of SOCS*.