Adaptively Exploiting d-Separators with Causal Bandits

Blair Bilodeau
(Joint work with Linbo Wang and Daniel M. Roy)
University of Toronto, Department of Statistical Sciences

November 15, 2022
The University of Chicago Rising Stars in Data Science Workshop

THE UNIVERSITY OF CHICAGO
DATA SCIENCE INSTITUTE
Rising Stars in Data Science

Motivation

Goal: Learn the intervention that has the largest positive causal effect on a variable of interest.

Motivation

Goal: Learn the intervention that has the largest positive causal effect on a variable of interest.

Impossible from observational data without unverifiable assumptions about the causal graph.

Motivation

Goal: Learn the intervention that has the largest positive causal effect on a variable of interest.

Impossible from observational data without unverifiable assumptions about the causal graph.

Instead, we can intervene...but this is expensive.

Motivation

Goal: Learn the intervention that has the largest positive causal effect on a variable of interest.

Impossible from observational data without unverifiable assumptions about the causal graph.

Instead, we can intervene...but this is expensive.

How can we most efficiently select which interventions to perform?

Motivation

Goal: Learn the intervention that has the largest positive causal effect on a variable of interest.

Impossible from observational data without unverifiable assumptions about the causal graph.

Instead, we can intervene...but this is expensive.

How can we most efficiently select which interventions to perform?

Existing causal algorithms:

Causal assumptions **hold** $\implies$ **More efficient interventions!**

Causal assumptions **fail** $\implies$ **Learn biased estimates.**

Motivation

Goal: Learn the intervention that has the largest positive causal effect on a variable of interest.

Impossible from observational data without unverifiable assumptions about the causal graph.

Instead, we can intervene...but this is expensive.

How can we most efficiently select which interventions to perform?

Existing causal algorithms:

Causal assumptions **hold** $\implies$ **More efficient interventions!**

Causal assumptions **fail** $\implies$ **Learn biased estimates.**

Our novel method:

Causal assumptions **hold** $\implies$ **Optimally efficient interventions!**

Causal assumptions **fail** $\implies$ **Still learn causal effects!**

Motivation

Goal: Learn the intervention that has the largest positive causal effect on a variable of interest.

Impossible from observational data without unverifiable assumptions about the causal graph.

Instead, we can intervene...but this is expensive.

How can we most efficiently select which interventions to perform?

Existing causal algorithms:

Causal assumptions **hold** $\implies$ **More efficient interventions!**

Causal assumptions **fail** $\implies$ **Learn biased estimates.**

Our novel method:

Causal assumptions **hold** $\implies$ **Optimally efficient interventions!**

Causal assumptions **fail** $\implies$ **Still learn causal effects!**

That is, our method *adapts* to the presence of causal structure.

Motivation

Goal: Learn the intervention that has the largest positive causal effect on a variable of interest.

Impossible from observational data without unverifiable assumptions about the causal graph.

Instead, we can intervene...but this is expensive.

How can we most efficiently select which interventions to perform?

Existing causal algorithms:

Causal assumptions **hold** $\implies$ **More efficient interventions!**

Causal assumptions **fail** $\implies$ **Learn biased estimates.**

Our novel method:

Causal assumptions **hold** $\implies$ **Optimally efficient interventions!**

Causal assumptions **fail** $\implies$ **Still learn causal effects!**

That is, our method *adapts* to the presence of causal structure.

Causal Inference with Interventions via Multi-Armed Bandits

Standard Multi-Armed Bandits

- Sequentially pick intervention $A_t \in \mathcal{A}$
- Observe reward Y_t
- Goal is to learn optimal intervention $\arg\max_{a \in \mathcal{A}} \mathbb{E}_a Y$

Causal Inference with Interventions via Multi-Armed Bandits

Standard Multi-Armed Bandits

- Sequentially pick intervention $A_t \in \mathcal{A}$
- Observe reward Y_t
- Goal is to learn optimal intervention $\arg\max_{a \in \mathcal{A}} \mathbb{E}_a Y$

Without more structure, this can be necessarily inefficient.

Causal Inference with Interventions via Multi-Armed Bandits

Standard Multi-Armed Bandits

- Sequentially pick intervention $A_t \in \mathcal{A}$
- Observe reward Y_t
- Goal is to learn optimal intervention $\arg\max_{a \in \mathcal{A}} \mathbb{E}_a Y$

Without more structure, this can be necessarily inefficient.

In practice, we also observe other information when we take an intervention.

Causal Inference with Interventions via Multi-Armed Bandits

Standard Multi-Armed Bandits

- Sequentially pick intervention $A_t \in \mathcal{A}$
- Observe reward Y_t
- Goal is to learn optimal intervention $\arg\max_{a \in \mathcal{A}} \mathbb{E}_a Y$

Without more structure, this can be necessarily inefficient.

In practice, we also observe other information when we take an intervention.

Multi-Armed Bandits with Post-Action Contexts: Also observe $Z_t \in \mathcal{Z}$.

We have no guarantees that observing Z_t will help us...but we would like to exploit it when we can.

Causal Inference with Interventions via Multi-Armed Bandits

Standard Multi-Armed Bandits

- Sequentially pick intervention $A_t \in \mathcal{A}$
- Observe reward Y_t
- Goal is to learn optimal intervention $\arg\max_{a \in \mathcal{A}} \mathbb{E}_a Y$

Without more structure, this can be necessarily inefficient.

In practice, we also observe other information when we take an intervention.

Multi-Armed Bandits with Post-Action Contexts: Also observe $Z_t \in \mathcal{Z}$.

We have no guarantees that observing Z_t will help us...but we would like to exploit it when we can.

An **environment** ν is a *fixed* collection of distributions on $(\mathcal{Z}, \mathcal{Y})$: one for each $a \in \mathcal{A}$.

Causal Inference with Interventions via Multi-Armed Bandits

Standard Multi-Armed Bandits

- Sequentially pick intervention $A_t \in \mathcal{A}$
- Observe reward Y_t
- Goal is to learn optimal intervention $\arg\max_{a \in \mathcal{A}} \mathbb{E}_a Y$

Without more structure, this can be necessarily inefficient.

In practice, we also observe other information when we take an intervention.

Multi-Armed Bandits with Post-Action Contexts: Also observe $Z_t \in \mathcal{Z}$.

We have no guarantees that observing Z_t will help us...but we would like to exploit it when we can.

An **environment** ν is a *fixed* collection of distributions on $(\mathcal{Z}, \mathcal{Y})$: one for each $a \in \mathcal{A}$.

A **policy** π maps the observed history $(A_1, Z_1, Y_1, \ldots, A_{t-1}, Z_{t-1}, Y_{t-1})$ to a distribution over A_t.

Causal Inference with Interventions via Multi-Armed Bandits

Standard Multi-Armed Bandits

- Sequentially pick intervention $A_t \in \mathcal{A}$
- Observe reward Y_t
- Goal is to learn optimal intervention $\arg\max_{a \in \mathcal{A}} \mathbb{E}_a Y$

Without more structure, this can be necessarily inefficient.

In practice, we also observe other information when we take an intervention.

Multi-Armed Bandits with Post-Action Contexts: Also observe $Z_t \in \mathcal{Z}$.

We have no guarantees that observing Z_t will help us...but we would like to exploit it when we can.

An **environment** ν is a *fixed* collection of distributions on $(\mathcal{Z}, \mathcal{Y})$: one for each $a \in \mathcal{A}$.

A **policy** π maps the observed history $(A_1, Z_1, Y_1, \ldots, A_{t-1}, Z_{t-1}, Y_{t-1})$ to a distribution over A_t.

Regret: $$R_{\nu,\pi}(T) = T \cdot \max_{a \in \mathcal{A}} \mathbb{E}_{\nu_a}\left[Y\right] - \mathbb{E}_{\nu,\pi}\left[\sum_{t=1}^{T} Y_t\right].$$

Standard Multi-Armed Bandits

- Sequentially pick intervention $A_t \in \mathcal{A}$
- Observe reward Y_t
- Goal is to learn optimal intervention $\arg\max_{a \in \mathcal{A}} \mathbb{E}_a Y$

Without more structure, this can be necessarily inefficient.

In practice, we also observe other information when we take an intervention.

Multi-Armed Bandits with Post-Action Contexts: Also observe $Z_t \in \mathcal{Z}$.

We have no guarantees that observing Z_t will help us...but we would like to exploit it when we can.

An **environment** ν is a *fixed* collection of distributions on $(\mathcal{Z}, \mathcal{Y})$: one for each $a \in \mathcal{A}$.

A **policy** π maps the observed history $(A_1, Z_1, Y_1, \ldots, A_{t-1}, Z_{t-1}, Y_{t-1})$ to a distribution over A_t.

Regret: $\quad R_{\nu,\pi}(T) = T \cdot \max_{a \in \mathcal{A}} \mathbb{E}_{\nu_a}\left[Y\right] - \mathbb{E}_{\nu,\pi}\left[\sum_{t=1}^{T} Y_t\right].$

The Punchline: High-Level Overview of Results

The Punchline: High-Level Overview of Results

We formalize when Z_t is helpful: conditionally benign environments.

The Punchline: High-Level Overview of Results

We formalize when Z_t is helpful: conditionally benign environments.

Existing causal algorithms have regret depending on $|\mathcal{Z}|$ instead of $|\mathcal{A}|$.

The Punchline: High-Level Overview of Results

We formalize when Z_t is helpful: conditionally benign environments.

Existing causal algorithms have regret depending on $|\mathcal{Z}|$ instead of $|\mathcal{A}|$.

Existing algorithms can have regret linear in T in the worst case.

This means they don't even have a consistent estimate of the causal effect!

The Punchline: High-Level Overview of Results

We formalize when Z_t is helpful: conditionally benign environments.

Existing causal algorithms have regret depending on $|\mathcal{Z}|$ instead of $|\mathcal{A}|$.

Existing algorithms can have regret linear in T in the worst case.

This means they don't even have a consistent estimate of the causal effect!

We formalize *adaptive minimax optimality* for the conditionally benign property.

Optimality is impossible: efficient interventions necessarily sacrifice worst-case performance.

The Punchline: High-Level Overview of Results

We formalize when Z_t is helpful: conditionally benign environments.

Existing causal algorithms have regret depending on $|\mathcal{Z}|$ instead of $|\mathcal{A}|$.

Existing algorithms can have regret linear in T in the worst case.

This means they don't even have a consistent estimate of the causal effect!

We formalize *adaptive minimax optimality* for the conditionally benign property.

Optimality is impossible: efficient interventions necessarily sacrifice worst-case performance.

We provide a new algorithm with:

a) **optimal performance for conditionally benign environments** and

b) **sublinear regret (always learns causal effects)**.

The Punchline: High-Level Overview of Results

We formalize when Z_t is helpful: conditionally benign environments.
Existing causal algorithms have regret depending on $|\mathcal{Z}|$ instead of $|\mathcal{A}|$.

Existing algorithms can have regret linear in T in the worst case.
This means they don't even have a consistent estimate of the causal effect!

We formalize *adaptive minimax optimality* for the conditionally benign property.
Optimality is impossible: efficient interventions necessarily sacrifice worst-case performance.

We provide a new algorithm with:

a) **optimal performance for conditionally benign environments** and
b) **sublinear regret (always learns causal effects)**.

Non-Adaptive Results

Non-Adaptive Results

Without any assumptions beyond IID,

UCB (Auer at al. 2002): $R_{\nu,\text{UCB}}(T) = \tilde{\Theta}(\sqrt{|\mathcal{A}|\,T})$

Non-Adaptive Results

Without any assumptions beyond IID,

UCB (Auer at al. 2002): $R_{\nu,\text{UCB}}(T) = \tilde{\Theta}(\sqrt{|\mathcal{A}|T})$

Definition (informal)

An environment ν is conditionally benign if and only if $\nu_a(Y \mid Z)$ is constant as a function of $a \in \mathcal{A}$.

Non-Adaptive Results

Without any assumptions beyond IID,

UCB (Auer at al. 2002): $R_{\nu,\text{UCB}}(T) = \tilde{\Theta}(\sqrt{|\mathcal{A}|\,T})$

Definition (informal)

An environment ν is conditionally benign if and only if $\nu_a(Y \mid Z)$ is constant as a function of $a \in \mathcal{A}$.

When the environment ν is conditionally benign and the marginal distributions $\nu(Z)$ are known,

C-UCB (Lu et al. 2020; BWR Thm 4.3): $R_{\nu,\text{C-UCB}}(T) = \tilde{\Theta}(\sqrt{|\mathcal{Z}|\,T})$

Non-Adaptive Results

Without any assumptions beyond IID,

UCB (Auer at al. 2002): $R_{\nu,\text{UCB}}(T) = \tilde{\Theta}(\sqrt{|\mathcal{A}|T})$

Definition (informal)

An environment ν is conditionally benign if and only if $\nu_a(Y \mid Z)$ is constant as a function of $a \in \mathcal{A}$.

When the environment ν is conditionally benign and the marginal distributions $\nu(Z)$ are known,

C-UCB (Lu et al. 2020; BWR Thm 4.3): $R_{\nu,\text{C-UCB}}(T) = \tilde{\Theta}(\sqrt{|\mathcal{Z}|T})$

Theorem: Existing algorithms do not adapt to failure of assumptions.

For every $\mathcal{A}$ and $\mathcal{Z}$, there exists ν such that

$$\lim_{T\to\infty} \frac{R_{\nu,\text{C-UCB}}(T)}{T} \geq 1/120.$$

Non-Adaptive Results

Without any assumptions beyond IID,
UCB (Auer at al. 2002): $R_{\nu,\text{UCB}}(T) = \tilde{\Theta}(\sqrt{|\mathcal{A}|T})$

Definition (informal)

An environment ν is conditionally benign if and only if $\nu_a(Y \mid Z)$ is constant as a function of $a \in \mathcal{A}$.

When the environment ν is conditionally benign and the marginal distributions $\nu(Z)$ are known,
C-UCB (Lu et al. 2020; BWR Thm 4.3): $R_{\nu,\text{C-UCB}}(T) = \tilde{\Theta}(\sqrt{|\mathcal{Z}|T})$

Theorem: Existing algorithms do not adapt to failure of assumptions.

For every $\mathcal{A}$ and $\mathcal{Z}$, there exists ν such that

$$\lim_{T\to\infty} \frac{R_{\nu,\text{C-UCB}}(T)}{T} \geq 1/120.$$

Can we adapt between these cases?

Non-Adaptive Results

Without any assumptions beyond IID,

UCB (Auer at al. 2002): $R_{\nu,\text{UCB}}(T) = \tilde{\Theta}(\sqrt{|\mathcal{A}|\,T})$

Definition (informal)

An environment ν is conditionally benign if and only if $\nu_a(Y \mid Z)$ is constant as a function of $a \in \mathcal{A}$.

When the environment ν is conditionally benign and the marginal distributions $\nu(Z)$ are known,

C-UCB (Lu et al. 2020; BWR Thm 4.3): $R_{\nu,\text{C-UCB}}(T) = \tilde{\Theta}(\sqrt{|\mathcal{Z}|\,T})$

Theorem: Existing algorithms do not adapt to failure of assumptions.

For every $\mathcal{A}$ and $\mathcal{Z}$, there exists ν such that

$$\lim_{T\to\infty} \frac{R_{\nu,\text{C-UCB}}(T)}{T} \geq 1/120.$$

Can we adapt between these cases?

Adaptive Results

Adaptive Results

Theorem: Strict adaptation to the conditionally benign property is impossible.

If π is such that $R_{\nu,\pi}(T) \leq O(\sqrt{|\mathcal{A}|\,T})$ for all ν,
there exists ν that is conditionally benign but $R_{\nu,\pi}(T) \geq \Omega(\sqrt{|\mathcal{A}|\,T})$.

Adaptive Results

Theorem: Strict adaptation to the conditionally benign property is impossible.

If π is such that $R_{\nu,\pi}(T) \leq O(\sqrt{|\mathcal{A}|\, T})$ for all ν,
there exists ν that is conditionally benign but $R_{\nu,\pi}(T) \geq \Omega(\sqrt{|\mathcal{A}|\, T})$.

Previous work requires that we know $\nu(Z) = \{\nu_a(Z) : a \in \mathcal{A}\}$ in advance.

Adaptive Results

Theorem: Strict adaptation to the conditionally benign property is impossible.

If π is such that $R_{\nu,\pi}(T) \leq O(\sqrt{|\mathcal{A}|\,T})$ for all ν,
there exists ν that is conditionally benign but $R_{\nu,\pi}(T) \geq \Omega(\sqrt{|\mathcal{A}|\,T})$.

Previous work requires that we know $\nu(Z) = \{\nu_a(Z) : a \in \mathcal{A}\}$ in advance.
Instead suppose that we have access to an estimate $\tilde{\nu}(Z)$.

Adaptive Results

Theorem: Strict adaptation to the conditionally benign property is impossible.

If π is such that $R_{\nu,\pi}(T) \leq O(\sqrt{|\mathcal{A}|\,T})$ for all ν,
there exists ν that is conditionally benign but $R_{\nu,\pi}(T) \geq \Omega(\sqrt{|\mathcal{A}|\,T})$.

Previous work requires that we know $\nu(Z) = \{\nu_a(Z) : a \in \mathcal{A}\}$ in advance.
Instead suppose that we have access to an estimate $\tilde{\nu}(Z)$.

Main Theorem: Our new algorithm HAC-UCB achieves non-trivial adaptivity.

For any $\mathcal{A}$, $\mathcal{Z}$, T, ν, and $\tilde{\nu}$,

$$R_{\nu,\text{HAC-UCB}}(T) \leq \tilde{O}(T^{3/4}).$$

Further, if ν is conditionally benign and $\|\nu(Z) - \tilde{\nu}(Z)\| \leq \varepsilon$,

$$R_{\nu,\text{HAC-UCB}}(T) \leq \tilde{O}(\sqrt{|\mathcal{Z}|T} + \varepsilon T)\,.$$

Adaptive Results

Theorem: Strict adaptation to the conditionally benign property is impossible.

If π is such that $R_{\nu,\pi}(T) \leq O(\sqrt{|\mathcal{A}|T})$ for all ν,
there exists ν that is conditionally benign but $R_{\nu,\pi}(T) \geq \Omega(\sqrt{|\mathcal{A}|T})$.

Previous work requires that we know $\nu(Z) = \{\nu_a(Z) : a \in \mathcal{A}\}$ in advance.
Instead suppose that we have access to an estimate $\tilde{\nu}(Z)$.

Main Theorem: Our new algorithm HAC-UCB achieves non-trivial adaptivity.

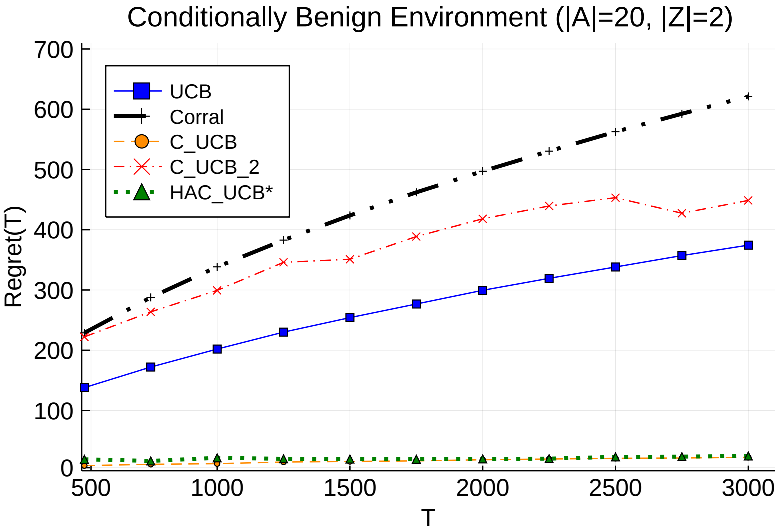
For any $\mathcal{A}$, $\mathcal{Z}$, T, ν, and $\tilde{\nu}$,

$$R_{\nu,\text{HAC-UCB}}(T) \leq \tilde{O}(T^{3/4}).$$

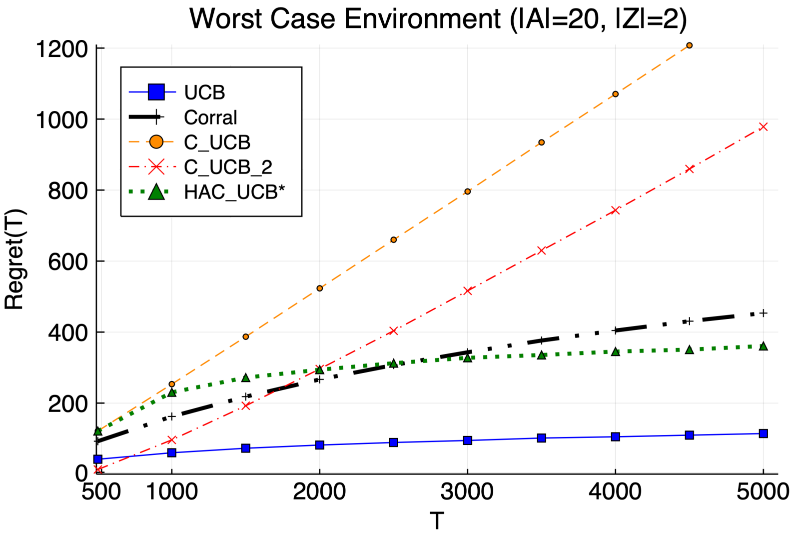
Further, if ν is conditionally benign and $\|\nu(Z) - \tilde{\nu}(Z)\| \leq \varepsilon$,

$$R_{\nu,\text{HAC-UCB}}(T) \leq \tilde{O}(\sqrt{|\mathcal{Z}|T} + \varepsilon T)\,.$$

Simulation Results



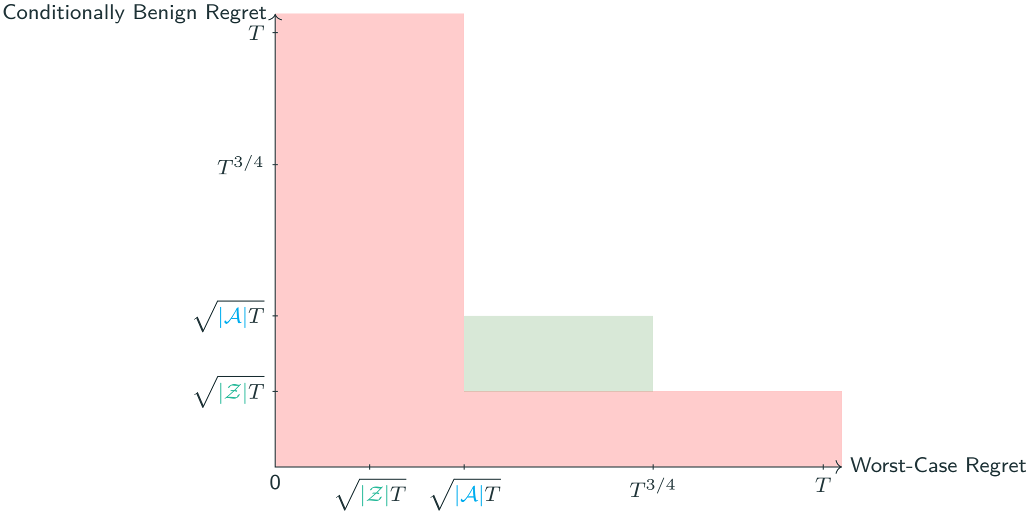
Simulation Results



Pareto Frontier of Causal Bandits

Worst-case optimal: UCB (Auer at al. 2002), Conditionally benign optimal: C-UCB (Lu et al. 2020)

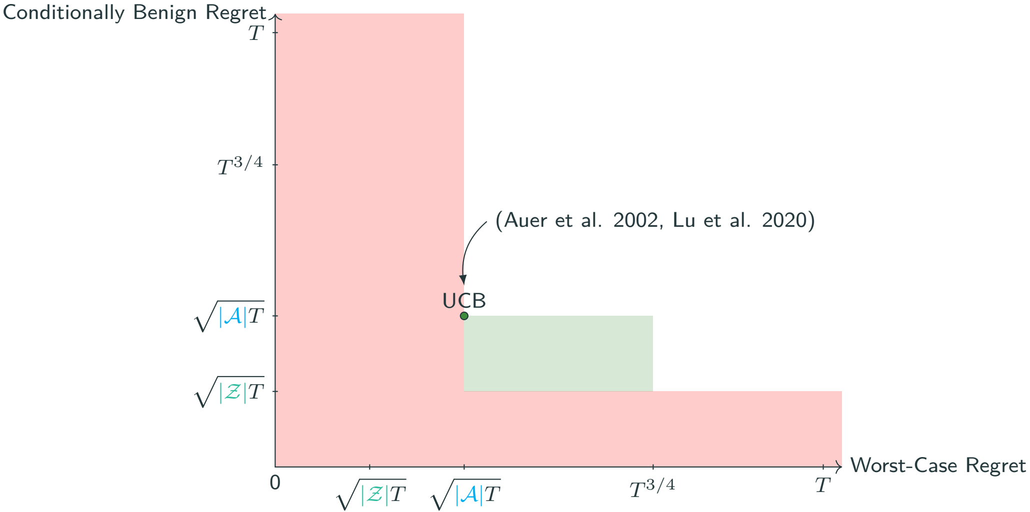
New algorithm: HAC-UCB (this work)



Pareto Frontier of Causal Bandits

Worst-case optimal: UCB (Auer at al. 2002), Conditionally benign optimal: C-UCB (Lu et al. 2020)

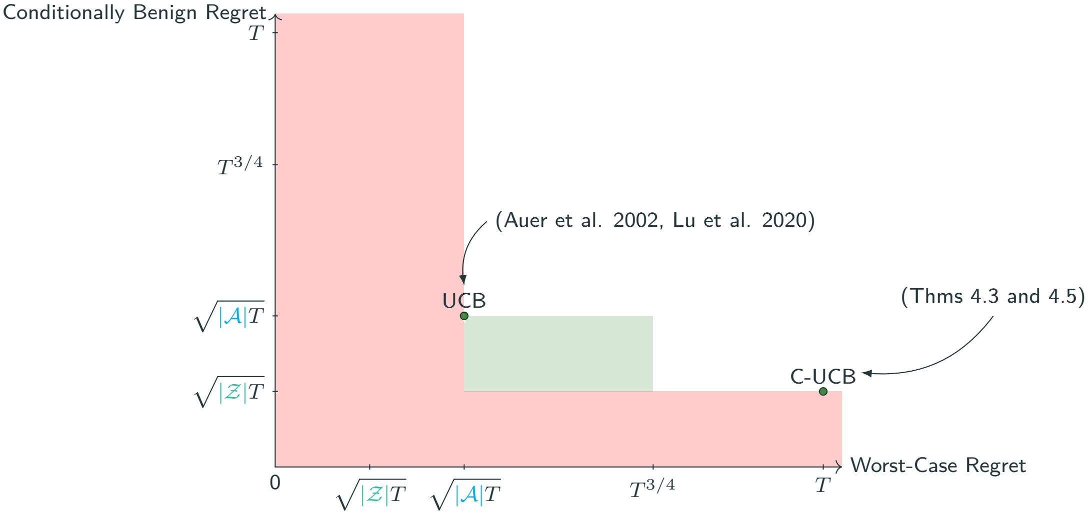
New algorithm: HAC-UCB (this work)



Pareto Frontier of Causal Bandits

Worst-case optimal: UCB (Auer at al. 2002), Conditionally benign optimal: C-UCB (Lu et al. 2020)

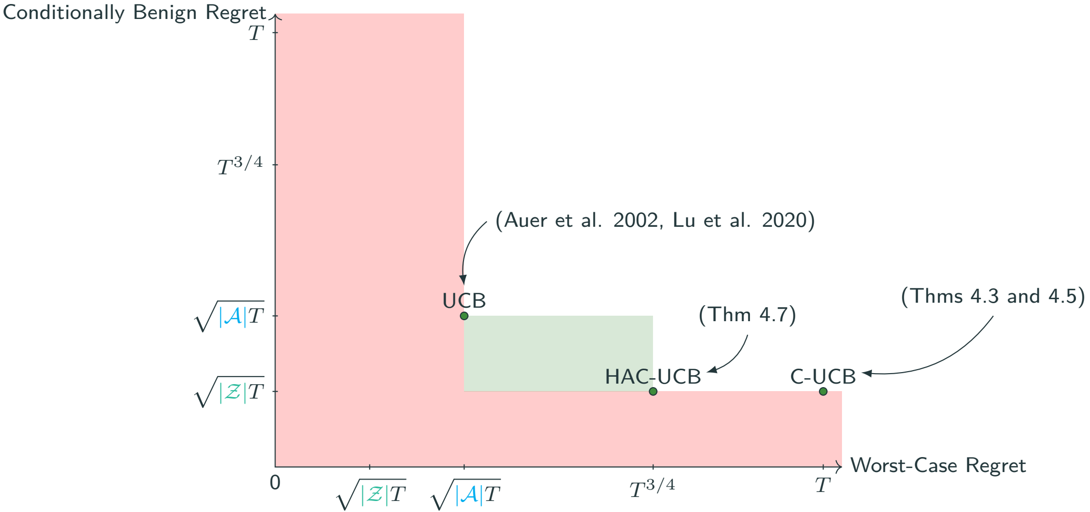
New algorithm: HAC-UCB (this work)



Pareto Frontier of Causal Bandits

Worst-case optimal: UCB (Auer at al. 2002), Conditionally benign optimal: C-UCB (Lu et al. 2020)

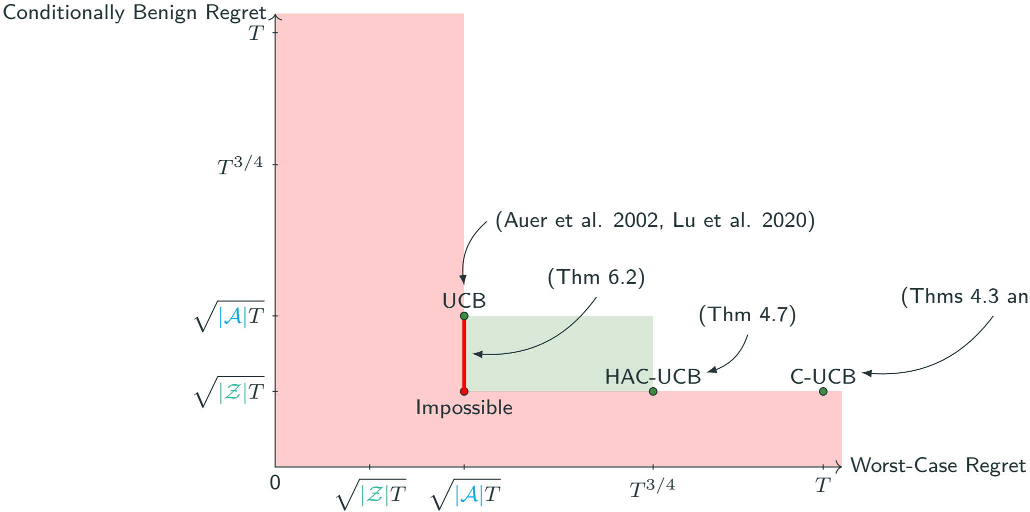
New algorithm: HAC-UCB (this work)



Pareto Frontier of Causal Bandits

Worst-case optimal: UCB (Auer at al. 2002), Conditionally benign optimal: C-UCB (Lu et al. 2020)

New algorithm: HAC-UCB (this work)



Understanding UCB and C-UCB

Understanding UCB and C-UCB

Upper Confidence Bound (UCB) Algorithm:

Understanding UCB and C-UCB

Upper Confidence Bound (UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(a)$ for each $t \in [T]$ and $a \in \mathcal{A}$

Understanding UCB and C-UCB

Upper Confidence Bound (UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(a)$ for each $t \in [T]$ and $a \in \mathcal{A}$
- Use concentration inequality to construct confidence bound $\text{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(T)/N_t(a)}$

Understanding UCB and C-UCB

Upper Confidence Bound (UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(a)$ for each $t \in [T]$ and $a \in \mathcal{A}$
- Use concentration inequality to construct confidence bound $\text{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(T)/N_t(a)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \text{UCB}_t(a)$

Understanding UCB and C-UCB

Upper Confidence Bound (UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(a)$ for each $t \in [T]$ and $a \in \mathcal{A}$
- Use concentration inequality to construct confidence bound $\mathrm{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(T)/N_t(a)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \mathrm{UCB}_t(a)$

Causal Upper Confidence Bound (C-UCB) Algorithm:

Understanding UCB and C-UCB

Upper Confidence Bound (UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(a)$ for each $t \in [T]$ and $a \in \mathcal{A}$
- Use concentration inequality to construct confidence bound $\text{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(T)/N_t(a)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \text{UCB}_t(a)$

Causal Upper Confidence Bound (C-UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(z)$ for each $t \in [T]$ and $z \in \mathcal{Z}$

Understanding UCB and C-UCB

Upper Confidence Bound (UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(a)$ for each $t \in [T]$ and $a \in \mathcal{A}$
- Use concentration inequality to construct confidence bound $\mathrm{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(T)/N_t(a)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \mathrm{UCB}_t(a)$

Causal Upper Confidence Bound (C-UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(z)$ for each $t \in [T]$ and $z \in \mathcal{Z}$
- Use concentration inequality to construct confidence bound $\mathrm{UCB}_t(z) = \hat{\mu}_t(z) + \sqrt{\log(T)/N_t(z)}$

Understanding UCB and C-UCB

Upper Confidence Bound (UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(a)$ for each $t \in [T]$ and $a \in \mathcal{A}$
- Use concentration inequality to construct confidence bound $\text{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(T)/N_t(a)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \text{UCB}_t(a)$

Causal Upper Confidence Bound (C-UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(z)$ for each $t \in [T]$ and $z \in \mathcal{Z}$
- Use concentration inequality to construct confidence bound $\text{UCB}_t(z) = \hat{\mu}_t(z) + \sqrt{\log(T)/N_t(z)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \sum_{z \in \mathcal{Z}} \text{UCB}_t(z) \mathbb{P}_{\tilde{\nu}_a}[Z = z]$

Understanding UCB and C-UCB

Upper Confidence Bound (UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(a)$ for each $t \in [T]$ and $a \in \mathcal{A}$
- Use concentration inequality to construct confidence bound $\text{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(T)/N_t(a)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \text{UCB}_t(a)$

Causal Upper Confidence Bound (C-UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(z)$ for each $t \in [T]$ and $z \in \mathcal{Z}$
- Use concentration inequality to construct confidence bound $\text{UCB}_t(z) = \hat{\mu}_t(z) + \sqrt{\log(T)/N_t(z)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \sum_{z \in \mathcal{Z}} \text{UCB}_t(z) \mathbb{P}_{\tilde{\nu}_a}[Z = z]$

Why does this work?

Understanding UCB and C-UCB

Upper Confidence Bound (UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(a)$ for each $t \in [T]$ and $a \in \mathcal{A}$
- Use concentration inequality to construct confidence bound $\text{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(T)/N_t(a)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \text{UCB}_t(a)$

Causal Upper Confidence Bound (C-UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(z)$ for each $t \in [T]$ and $z \in \mathcal{Z}$
- Use concentration inequality to construct confidence bound $\text{UCB}_t(z) = \hat{\mu}_t(z) + \sqrt{\log(T)/N_t(z)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \sum_{z \in \mathcal{Z}} \text{UCB}_t(z) \mathbb{P}_{\tilde{\nu}_a}[Z = z]$

Why does this work?

If all parents are observed (more generally, ν is conditionally benign) and $\tilde{\nu}(Z)$ is accurate,

$$\sum_{z \in \mathcal{Z}} \text{UCB}_t(z) \mathbb{P}_{\tilde{\nu}_a}[Z = z] \approx \text{UCB}_t(a),$$

but concentration only requires a union bound of size $|\mathcal{Z}|$ instead of size $|\mathcal{A}|$.

Understanding UCB and C-UCB

Upper Confidence Bound (UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(a)$ for each $t \in [T]$ and $a \in \mathcal{A}$
- Use concentration inequality to construct confidence bound $\text{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(T)/N_t(a)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \text{UCB}_t(a)$

Causal Upper Confidence Bound (C-UCB) Algorithm:

- Maintain empirical mean estimate $\hat{\mu}_t(z)$ for each $t \in [T]$ and $z \in \mathcal{Z}$
- Use concentration inequality to construct confidence bound $\text{UCB}_t(z) = \hat{\mu}_t(z) + \sqrt{\log(T)/N_t(z)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \sum_{z \in \mathcal{Z}} \text{UCB}_t(z)\mathbb{P}_{\tilde{\nu}_a}[Z = z]$

Why does this work?

If all parents are observed (more generally, ν is conditionally benign) and $\tilde{\nu}(Z)$ is accurate,

$$\sum_{z \in \mathcal{Z}} \text{UCB}_t(z)\mathbb{P}_{\tilde{\nu}_a}[Z = z] \approx \text{UCB}_t(a),$$

but concentration only requires a union bound of size $|\mathcal{Z}|$ instead of size $|\mathcal{A}|$.

Adapting with Hypothesis Testing: HAC-UCB

Adapting with Hypothesis Testing: HAC-UCB

Intuition: Optimistically play C-UCB until a hypothesis test for conditionally benign fails, then play UCB.

Adapting with Hypothesis Testing: HAC-UCB

Intuition: Optimistically play C-UCB until a hypothesis test for conditionally benign fails, then play UCB.

(1) Initial Exploration

Adapting with Hypothesis Testing: HAC-UCB

Intuition: Optimistically play C-UCB until a hypothesis test for conditionally benign fails, then play UCB.

(1) Initial Exploration

Uniformly sample $a \in \mathcal{A}$ for $\sqrt{T}/|\mathcal{A}|$ rounds.

Compute MLE estimate $\hat{\nu}$ of $(\nu_a(Z))_{a\in\mathcal{A}}$. If $\sup_{a\in\mathcal{A}} \|\tilde{\nu}_a - \hat{\nu}_a\|_1 \gtrsim T^{-1/4}$, set $\tilde{\nu} \leftarrow \hat{\nu}$.

Adapting with Hypothesis Testing: HAC-UCB

Intuition: Optimistically play C-UCB until a hypothesis test for conditionally benign fails, then play UCB.

(1) Initial Exploration

Uniformly sample $a \in \mathcal{A}$ for $\sqrt{T}/|\mathcal{A}|$ rounds.

Compute MLE estimate $\hat{\nu}$ of $(\nu_a(Z))_{a\in\mathcal{A}}$. If $\sup_{a\in\mathcal{A}} \|\tilde{\nu}_a - \hat{\nu}_a\|_1 \gtrsim T^{-1/4}$, set $\tilde{\nu} \leftarrow \hat{\nu}$.

Optimistic Phase: For each round t...

$\mathrm{UCB}_t(a) \approx \hat{\mathbb{E}}_{\nu_a}[Y] + \sqrt{(\log T)/N_a(t)}$.

$\widetilde{\mathrm{UCB}}_t(a) \approx \sum_{z\in\mathcal{Z}} [\hat{\mathbb{E}}_\nu[Y \mid Z = z] + \sqrt{(\log T)/N_z(t)}]\tilde{\nu}_a(Z = z)$.

If $\mathrm{UCB}_t(a) \approx \widetilde{\mathrm{UCB}}_t(a)$, play $A_{t+1} = \arg\max_{a\in\mathcal{A}} \widetilde{\mathrm{UCB}}_t(a)$.

Otherwise, switch to Pessimistic Phase.

Adapting with Hypothesis Testing: HAC-UCB

Intuition: Optimistically play C-UCB until a hypothesis test for conditionally benign fails, then play UCB.

(1) Initial Exploration

Uniformly sample $a \in \mathcal{A}$ for $\sqrt{T}/|\mathcal{A}|$ rounds.

Compute MLE estimate $\hat{\nu}$ of $(\nu_a(Z))_{a \in \mathcal{A}}$. If $\sup_{a \in \mathcal{A}} \|\tilde{\nu}_a - \hat{\nu}_a\|_1 \gtrsim T^{-1/4}$, set $\tilde{\nu} \leftarrow \hat{\nu}$.

Optimistic Phase: For each round t...

$\mathrm{UCB}_t(a) \approx \hat{\mathbb{E}}_{\nu_a}[Y] + \sqrt{(\log T)/N_a(t)}$.

$\widetilde{\mathrm{UCB}}_t(a) \approx \sum_{z \in \mathcal{Z}} [\hat{\mathbb{E}}_{\nu}[Y \mid Z = z] + \sqrt{(\log T)/N_z(t)}]\tilde{\nu}_a(Z = z)$.

If $\mathrm{UCB}_t(a) \approx \widetilde{\mathrm{UCB}}_t(a)$, play $A_{t+1} = \arg\max_{a \in \mathcal{A}} \widetilde{\mathrm{UCB}}_t(a)$.

Otherwise, switch to Pessimistic Phase.

Pessimistic Phase: For remaining rounds t, play $A_{t+1} = \arg\max_{a \in \mathcal{A}} \mathrm{UCB}_t(a)$.

Adapting with Hypothesis Testing: HAC-UCB

Intuition: Optimistically play C-UCB until a hypothesis test for conditionally benign fails, then play UCB.

(1) Initial Exploration

Uniformly sample $a \in \mathcal{A}$ for $\sqrt{T}/|\mathcal{A}|$ rounds.

Compute MLE estimate $\hat{\nu}$ of $(\nu_a(Z))_{a\in\mathcal{A}}$. If $\sup_{a\in\mathcal{A}} \|\tilde{\nu}_a - \hat{\nu}_a\|_1 \gtrsim T^{-1/4}$, set $\tilde{\nu} \leftarrow \hat{\nu}$.

Optimistic Phase: For each round t...

$\mathrm{UCB}_t(a) \approx \hat{\mathbb{E}}_{\nu_a}[Y] + \sqrt{(\log T)/N_a(t)}$.

$\widetilde{\mathrm{UCB}}_t(a) \approx \sum_{z\in\mathcal{Z}} [\hat{\mathbb{E}}_{\nu}[Y \mid Z = z] + \sqrt{(\log T)/N_z(t)}]\tilde{\nu}_a(Z = z)$.

If $\mathrm{UCB}_t(a) \approx \widetilde{\mathrm{UCB}}_t(a)$, play $A_{t+1} = \arg\max_{a\in\mathcal{A}} \widetilde{\mathrm{UCB}}_t(a)$.

Otherwise, switch to Pessimistic Phase.

Pessimistic Phase: For remaining rounds t, play $A_{t+1} = \arg\max_{a\in\mathcal{A}} \mathrm{UCB}_t(a)$.

Proof Sketch for HAC-UCB

Proof Sketch for HAC-UCB

(1) Exploration Rounds

Proof Sketch for HAC-UCB

(1) Exploration Rounds

In the worst case, C-UCB never plays the optimal $a \in \mathcal{A}$.

Proof Sketch for HAC-UCB

(1) Exploration Rounds

In the worst case, C-UCB never plays the optimal $a \in \mathcal{A}$.

To circumvent this, we explore each $a \in \mathcal{A}$ for an initial $\sqrt{T}/|\mathcal{A}|$ rounds.

Proof Sketch for HAC-UCB

(1) Exploration Rounds

In the worst case, C-UCB never plays the optimal $a \in \mathcal{A}$.

To circumvent this, we explore each $a \in \mathcal{A}$ for an initial $\sqrt{T}/|\mathcal{A}|$ rounds.

This is fine from a minimax perspective since even conditionally benign forces $\sqrt{T}$ regret.

Proof Sketch for HAC-UCB

(1) Exploration Rounds

In the worst case, C-UCB never plays the optimal $a \in \mathcal{A}$.

To circumvent this, we explore each $a \in \mathcal{A}$ for an initial $\sqrt{T}/|\mathcal{A}|$ rounds.

This is fine from a minimax perspective since even conditionally benign forces $\sqrt{T}$ regret.

Estimating a multinomial to scale ε takes $\approx 1/\varepsilon^2$ samples,
so we also use the initial exploration to obtain an $\varepsilon = T^{-1/4}$ accurate estimate of $\nu(\mathcal{Z})$.

Proof Sketch for HAC-UCB

(1) Exploration Rounds

In the worst case, C-UCB never plays the optimal $a \in \mathcal{A}$.

To circumvent this, we explore each $a \in \mathcal{A}$ for an initial $\sqrt{T}/|\mathcal{A}|$ rounds.

This is fine from a minimax perspective since even conditionally benign forces $\sqrt{T}$ regret.

Estimating a multinomial to scale ε takes $\approx 1/\varepsilon^2$ samples,
so we also use the initial exploration to obtain an $\varepsilon = T^{-1/4}$ accurate estimate of $\nu(\mathcal{Z})$.

(2) Optimistic Rounds

Proof Sketch for HAC-UCB

(1) Exploration Rounds

In the worst case, C-UCB never plays the optimal $a \in \mathcal{A}$.
To circumvent this, we explore each $a \in \mathcal{A}$ for an initial $\sqrt{T}/|\mathcal{A}|$ rounds.
This is fine from a minimax perspective since even conditionally benign forces $\sqrt{T}$ regret.
Estimating a multinomial to scale ε takes $\approx 1/\varepsilon^2$ samples,
so we also use the initial exploration to obtain an $\varepsilon = T^{-1/4}$ accurate estimate of $\nu(\mathcal{Z})$.

(2) Optimistic Rounds

a) If the conditionally benign assumption holds,
$\mathrm{UCB}_t(a) \approx \widetilde{\mathrm{UCB}}_t(a)$ and the algorithm correctly plays optimistically.

b) If the conditionally benign assumption fails,
either $\mathrm{UCB}_t(a) \not\approx \widetilde{\mathrm{UCB}}_t(a)$ and the algorithm correctly plays pessimistically,
or the regret incurred from playing optimistically is still sufficiently small.

Proof Sketch for HAC-UCB

(1) Exploration Rounds

In the worst case, C-UCB never plays the optimal $a \in \mathcal{A}$.

To circumvent this, we explore each $a \in \mathcal{A}$ for an initial $\sqrt{T}/|\mathcal{A}|$ rounds.

This is fine from a minimax perspective since even conditionally benign forces $\sqrt{T}$ regret.

Estimating a multinomial to scale ε takes $\approx 1/\varepsilon^2$ samples,
so we also use the initial exploration to obtain an $\varepsilon = T^{-1/4}$ accurate estimate of $\nu(\mathcal{Z})$.

(2) Optimistic Rounds

a) If the conditionally benign assumption holds,
$\mathrm{UCB}_t(a) \approx \widetilde{\mathrm{UCB}}_t(a)$ and the algorithm correctly plays optimistically.

b) If the conditionally benign assumption fails,
either $\mathrm{UCB}_t(a) \not\approx \widetilde{\mathrm{UCB}}_t(a)$ and the algorithm correctly plays pessimistically,
or the regret incurred from playing optimistically is still sufficiently small.

More on Conditionally Benign

More on Conditionally Benign

Suppose we have a fixed DAG $\mathcal{G}$ on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$.

More on Conditionally Benign

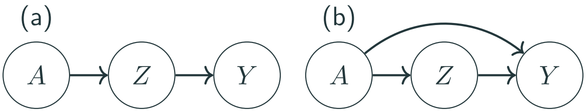
Suppose we have a fixed DAG $\mathcal{G}$ on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$.

(a)

(a) conditionally benign and d-separated

More on Conditionally Benign

Suppose we have a fixed DAG $\mathcal{G}$ on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$.

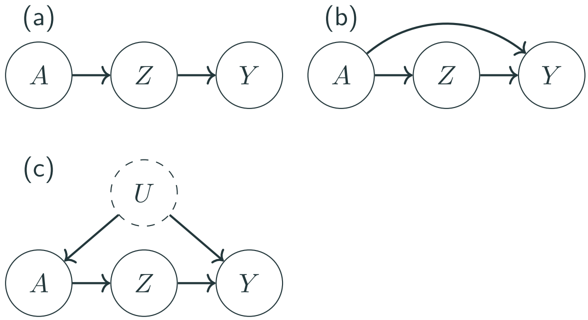


(a) conditionally benign and d-separated

(b) not conditionally benign

More on Conditionally Benign

Suppose we have a fixed DAG $\mathcal{G}$ on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$.



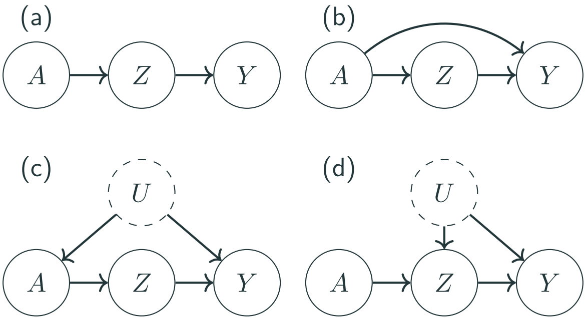
(a) conditionally benign and d-separated

(b) not conditionally benign

(c) conditionally benign through front-door, not d-separated

More on Conditionally Benign

Suppose we have a fixed DAG $\mathcal{G}$ on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$.



(a) conditionally benign and d-separated

(b) not conditionally benign

(c) conditionally benign through front-door, not d-separated

(d) no adjustment possible, not conditionally benign

More on Conditionally Benign

More on Conditionally Benign

Suppose we have a fixed DAG $\mathcal{G}$ on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$.

More on Conditionally Benign

Suppose we have a fixed DAG $\mathcal{G}$ on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$.

Theorem

Let $\mathcal{A}$ be all hard interventions.

Z d-separates Y from A on $\mathcal{G}$ if and only if every Markov relative ν on $\mathcal{G}$ is conditionally benign on $\mathcal{A}$.

More on Conditionally Benign

Suppose we have a fixed DAG $\mathcal{G}$ on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$. Let $\mathcal{G}_{\overline{A}}$ denote the graph with edges into A removed.

Theorem

Let $\mathcal{A}$ be all hard interventions.

Z d-separates Y from A on $\mathcal{G}$ if and only if every Markov relative ν on $\mathcal{G}$ is conditionally benign on $\mathcal{A}$.

More on Conditionally Benign

Suppose we have a fixed DAG $\mathcal{G}$ on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$. Let $\mathcal{G}_{\overline{A}}$ denote the graph with edges into A removed.

Theorem

Let $\mathcal{A}$ be all hard interventions.

Z d-separates Y from A on $\mathcal{G}$ if and only if every Markov relative ν on $\mathcal{G}$ is conditionally benign on $\mathcal{A}$.

Theorem

Let $\mathcal{A}_0$ be all hard interventions except the null (observational) intervention.

Z d-separates Y from A on $\mathcal{G}_{\overline{A}}$ if and only if every Markov relative ν on $\mathcal{G}$ is conditionally benign on $\mathcal{A}_0$.

More on Conditionally Benign

Suppose we have a fixed DAG $\mathcal{G}$ on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$. Let $\mathcal{G}_{\overline{A}}$ denote the graph with edges into A removed.

Theorem

Let $\mathcal{A}$ be all hard interventions.

Z d-separates Y from A on $\mathcal{G}$ if and only if every Markov relative ν on $\mathcal{G}$ is conditionally benign on $\mathcal{A}$.

Theorem

Let $\mathcal{A}_0$ be all hard interventions except the null (observational) intervention.

Z d-separates Y from A on $\mathcal{G}_{\overline{A}}$ if and only if every Markov relative ν on $\mathcal{G}$ is conditionally benign on $\mathcal{A}_0$.

Proposition

If Z satisfies the front-door criterion with respect to (A, Y) on $\mathcal{G}$ then Z d-separates Y from A on $\mathcal{G}_{\overline{A}}$.