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

March 29, 2022

Presented to the Approximately Correct Machine Intelligence Lab

Motivation

Motivation

Assumptions are used to develop statistical methods and provide guarantees,

Motivation

Assumptions are used to develop statistical methods and provide guarantees,
leaving us susceptible to sharply degrading performance under failure of assumptions.

Motivation

Assumptions are used to develop statistical methods and provide guarantees,
leaving us susceptible to sharply degrading performance under failure of assumptions.

Want to act optimally without having to know how data are generated.

Motivation

Assumptions are used to develop statistical methods and provide guarantees,
leaving us susceptible to sharply degrading performance under failure of assumptions.

Want to act optimally without having to know how data are generated.

Intuitively, we might expect that...

Motivation

Assumptions are used to develop statistical methods and provide guarantees, leaving us susceptible to sharply degrading performance under failure of assumptions.

Want to act optimally without having to know how data are generated.

Intuitively, we might expect that...

Better performance is possible in certain *benign* settings, and

Motivation

Assumptions are used to develop statistical methods and provide guarantees,
leaving us susceptible to sharply degrading performance under failure of assumptions.

Want to act optimally without having to know how data are generated.

Intuitively, we might expect that...

Better performance is possible in certain *benign* settings, and
we can learn whether the setting is benign from the data.

Motivation

Assumptions are used to develop statistical methods and provide guarantees,
leaving us susceptible to sharply degrading performance under failure of assumptions.

Want to act optimally without having to know how data are generated.

Intuitively, we might expect that...

Better performance is possible in certain *benign* settings, and
we can learn whether the setting is benign from the data.

Can we design **robust** decision methods that **adapt** to the failure of those assumptions?

Motivation

Assumptions are used to develop statistical methods and provide guarantees,
leaving us susceptible to sharply degrading performance under failure of assumptions.

Want to act optimally without having to know how data are generated.

Intuitively, we might expect that...

Better performance is possible in certain *benign* settings, and
we can learn whether the setting is benign from the data.

Can we design **robust** decision methods that **adapt** to the failure of those assumptions?

Adapting means we simultaneously...

Motivation

Assumptions are used to develop statistical methods and provide guarantees,
leaving us susceptible to sharply degrading performance under failure of assumptions.

Want to act optimally without having to know how data are generated.

Intuitively, we might expect that...

Better performance is possible in certain *benign* settings, and
we can learn whether the setting is benign from the data.

Can we design **robust** decision methods that **adapt** to the failure of those assumptions?

Adapting means we simultaneously...

...benefit from assumptions when they hold,

Motivation

Assumptions are used to develop statistical methods and provide guarantees,
leaving us susceptible to sharply degrading performance under failure of assumptions.

Want to act optimally without having to know how data are generated.

Intuitively, we might expect that...

Better performance is possible in certain *benign* settings, and
we can learn whether the setting is benign from the data.

Can we design **robust** decision methods that **adapt** to the failure of those assumptions?

Adapting means we simultaneously...

...benefit from assumptions when they hold,
...but still do "as well as possible" when they fail,

Motivation

Assumptions are used to develop statistical methods and provide guarantees,
leaving us susceptible to sharply degrading performance under failure of assumptions.

Want to act optimally without having to know how data are generated.

Intuitively, we might expect that...

Better performance is possible in certain *benign* settings, and
we can learn whether the setting is benign from the data.

Can we design **robust** decision methods that **adapt** to the failure of those assumptions?

Adapting means we simultaneously...

...benefit from assumptions when they hold,
...but still do "as well as possible" when they fail,
...without knowing which case we are in.

Motivation

Assumptions are used to develop statistical methods and provide guarantees,
leaving us susceptible to sharply degrading performance under failure of assumptions.

Want to act optimally without having to know how data are generated.

Intuitively, we might expect that...

Better performance is possible in certain *benign* settings, and
we can learn whether the setting is benign from the data.

Can we design **robust** decision methods that **adapt** to the failure of those assumptions?

Adapting means we simultaneously...

...benefit from assumptions when they hold,
...but still do "as well as possible" when they fail,
...without knowing which case we are in.

Isn’t this impossible?

Isn't this impossible?

E.g., we can't learn the best intervention from observational data without assumptions.

Isn't this impossible?

E.g., we can't learn the best intervention from observational data without assumptions.

To circumvent impossibility, we can change the rules of the game or how we measure performance.

Isn't this impossible?

E.g., we can't learn the best intervention from observational data without assumptions.

To circumvent impossibility, we can change the rules of the game or how we measure performance.

Some Examples

Isn't this impossible?

E.g., we can't learn the best intervention from observational data without assumptions.

To circumvent impossibility, we can change the rules of the game or how we measure performance.

Some Examples

Excess risk once we remove the well-specified assumption.

Isn't this impossible?

E.g., we can't learn the best intervention from observational data without assumptions.

To circumvent impossibility, we can change the rules of the game or how we measure performance.

Some Examples

- Excess risk once we remove the well-specified assumption.
- Regret and sequential decisions once we remove the i.i.d. assumption.

Isn't this impossible?

E.g., we can't learn the best intervention from observational data without assumptions.

To circumvent impossibility, we can change the rules of the game or how we measure performance.

Some Examples

- Excess risk once we remove the well-specified assumption.
- Regret and sequential decisions once we remove the i.i.d. assumption.
- Active learning queries once we remove assumptions about distribution shift.

Isn't this impossible?

E.g., we can't learn the best intervention from observational data without assumptions.

To circumvent impossibility, we can change the rules of the game or how we measure performance.

Some Examples

Excess risk once we remove the well-specified assumption.

Regret and sequential decisions once we remove the i.i.d. assumption.

Active learning queries once we remove assumptions about distribution shift.

This work: Allowing interventions once we remove assumptions about confounders.

Isn't this impossible?

E.g., we can't learn the best intervention from observational data without assumptions.

To circumvent impossibility, we can change the rules of the game or how we measure performance.

Some Examples

- Excess risk once we remove the well-specified assumption.
- Regret and sequential decisions once we remove the i.i.d. assumption.
- Active learning queries once we remove assumptions about distribution shift.

This work: Allowing interventions once we remove assumptions about confounders.

Why assumptions on confounding?

Why assumptions on confounding?

Goal: Identify which intervention maximizes reward.

Why assumptions on confounding?

Goal: Identify which intervention maximizes reward.

For observational data, we **must** make assumptions that answer...

Why assumptions on confounding?

Goal: Identify which intervention maximizes reward.

For observational data, we **must** make assumptions that answer...

How do unobserved variables confound the relationship between intervention and outcome?

Why assumptions on confounding?

Goal: Identify which intervention maximizes reward.

For observational data, we **must** make assumptions that answer...

How do unobserved variables confound the relationship between intervention and outcome?

Prior work: Assumptions about confounding also enable us to select interventions more efficiently.

Why assumptions on confounding?

Goal: Identify which intervention maximizes reward.

For observational data, we **must** make assumptions that answer...

How do unobserved variables confound the relationship between intervention and outcome?

Prior work: Assumptions about confounding also enable us to select interventions more efficiently.

Since we no longer require such assumptions to achieve non-trivial performance...

Why assumptions on confounding?

Goal: Identify which intervention maximizes reward.

For observational data, we **must** make assumptions that answer...

How do unobserved variables confound the relationship between intervention and outcome?

Prior work: Assumptions about confounding also enable us to select interventions more efficiently.

Since we no longer require such assumptions to achieve non-trivial performance...
...it is natural to ask if we can adapt to their presence.

Why assumptions on confounding?

Goal: Identify which intervention maximizes reward.

For observational data, we **must** make assumptions that answer...

How do unobserved variables confound the relationship between intervention and outcome?

Prior work: Assumptions about confounding also enable us to select interventions more efficiently.

Since we no longer require such assumptions to achieve non-trivial performance...
...it is natural to ask if we can adapt to their presence.

With interventions, non-trivial adaptivity is possible without *any assumptions on confounding*.

Why assumptions on confounding?

Goal: Identify which intervention maximizes reward.

For observational data, we **must** make assumptions that answer...

How do unobserved variables confound the relationship between intervention and outcome?

Prior work: Assumptions about confounding also enable us to select interventions more efficiently.

Since we no longer require such assumptions to achieve non-trivial performance...
...it is natural to ask if we can adapt to their presence.

With interventions, non-trivial <u>adaptivity</u> is possible without *any assumptions on confounding*.

How do we measure performance?

How do we measure performance?

We measure performance with **regret**, which compares

the cumulative reward of our policy to the cumulative reward of the best fixed intervention.

How do we measure performance?

We measure performance with **regret**, which compares

the cumulative reward of our policy to the cumulative reward of the best fixed intervention.

Why compete against the best fixed intervention?

How do we measure performance?

We measure performance with **regret**, which compares

the cumulative reward of our policy to the cumulative reward of the best fixed intervention.

Why compete against the best fixed intervention?

We cannot hope to achieve good objective performance (e.g., risk vs excess risk).

How do we measure performance?

We measure performance with **regret**, which compares

the cumulative reward of our policy to the cumulative reward of the best fixed intervention.

Why compete against the best fixed intervention?

We cannot hope to achieve good objective performance (e.g., risk vs excess risk).

One reason to identify treatment effects is to identify the best treatment.

How do we measure performance?

We measure performance with **regret**, which compares

the cumulative reward of our policy to the cumulative reward of the best fixed intervention.

Why compete against the best fixed intervention?

We cannot hope to achieve good objective performance (e.g., risk vs excess risk).

One reason to identify treatment effects is to identify the best treatment.

We ask: "How well did we do compared to if someone told us the best treatment in advance?"

How do we measure performance?

We measure performance with **regret**, which compares

the cumulative reward of our policy to the cumulative reward of the best fixed intervention.

Why compete against the best fixed intervention?

We cannot hope to achieve good objective performance (e.g., risk vs excess risk).

One reason to identify treatment effects is to identify the best treatment.

We ask: "How well did we do compared to if someone told us the best treatment in advance?"

Why measure performance cumulatively?

How do we measure performance?

We measure performance with **regret**, which compares

the cumulative reward of our policy to the cumulative reward of the best fixed intervention.

Why compete against the best fixed intervention?

We cannot hope to achieve good objective performance (e.g., risk vs excess risk).

One reason to identify treatment effects is to identify the best treatment.

We ask: "How well did we do compared to if someone told us the best treatment in advance?"

Why measure performance cumulatively?

Best-arm identification outputs a single intervention after T rounds.

How do we measure performance?

We measure performance with **regret**, which compares

the cumulative reward of our policy to the cumulative reward of the best fixed intervention.

Why compete against the best fixed intervention?

We cannot hope to achieve good objective performance (e.g., risk vs excess risk).

One reason to identify treatment effects is to identify the best treatment.

We ask: "How well did we do compared to if someone told us the best treatment in advance?"

Why measure performance cumulatively?

Best-arm identification outputs a single intervention after T rounds.

We want to (a) identify a good intervention and (b) do so quickly...

How do we measure performance?

We measure performance with **regret**, which compares

the cumulative reward of our policy to the cumulative reward of the best fixed intervention.

Why compete against the best fixed intervention?

We cannot hope to achieve good objective performance (e.g., risk vs excess risk).

One reason to identify treatment effects is to identify the best treatment.

We ask: "How well did we do compared to if someone told us the best treatment in advance?"

Why measure performance cumulatively?

Best-arm identification outputs a single intervention after T rounds.

We want to (a) identify a good intervention and (b) do so quickly...

...while *also* (c) making our interventions as harmless as possible.

How do we measure performance?

We measure performance with **regret**, which compares

the cumulative reward of our policy to the cumulative reward of the best fixed intervention.

Why compete against the best fixed intervention?

We cannot hope to achieve good objective performance (e.g., risk vs excess risk).

One reason to identify treatment effects is to identify the best treatment.

We ask: "How well did we do compared to if someone told us the best treatment in advance?"

Why measure performance cumulatively?

Best-arm identification outputs a single intervention after T rounds.

We want to (a) identify a good intervention and (b) do so quickly...

...while *also* (c) making our interventions as harmless as possible.

Let's formalize the setting we're working in.

Let's formalize the setting we're working in.

Multi-Armed Bandits

Multi-Armed Bandits

Fix a set of allowable *actions* $\mathcal{A}$.

Causal Inference with Interventions via Multi-Armed Bandits

Multi-Armed Bandits

Fix a set of allowable *actions* $\mathcal{A}$.

For each interaction $t \in [T]$:

Causal Inference with Interventions via Multi-Armed Bandits

Multi-Armed Bandits

Fix a set of allowable *actions* $\mathcal{A}$.

For each interaction $t \in [T]$:

- Player selects $A_t \in \mathcal{A}$

Causal Inference with Interventions via Multi-Armed Bandits

Multi-Armed Bandits

Fix a set of allowable *actions* $\mathcal{A}$.

For each interaction $t \in [T]$:

- Player selects $A_t \in \mathcal{A}$
- Environment generates $Y_t \in [0, 1]$

Multi-Armed Bandits

Fix a set of allowable *actions* $\mathcal{A}$.

For each interaction $t \in [T]$:

- Player selects $A_t \in \mathcal{A}$
- Environment generates $Y_t \in [0, 1]$
- Player observes Y_t

Causal Inference with Interventions via Multi-Armed Bandits

Multi-Armed Bandits

Fix a set of allowable *actions* $\mathcal{A}$.

For each interaction $t \in [T]$:

- Player selects $A_t \in \mathcal{A}$
- Environment generates $Y_t \in [0, 1]$
- Player observes Y_t

We do not observe the counterfactual:

How would the environment have generated Y_t if the player picked a different A_t?

Multi-Armed Bandits

Fix a set of allowable *actions* $\mathcal{A}$.

For each interaction $t \in [T]$:

- Player selects $A_t \in \mathcal{A}$
- Environment generates $Y_t \in [0, 1]$
- Player observes Y_t

We do not observe the counterfactual:

How would the environment have generated Y_t if the player picked a different A_t?

Suppose we also observe other variables that might help us learn this.

Multi-Armed Bandits with Post-Action Contexts

Fix a set of allowable *actions* $\mathcal{A}$ and *post-action contexts* $\mathcal{Z}$.

For each interaction $t \in [T]$:

- Player selects $A_t \in \mathcal{A}$
- Environment generates $Y_t \in [0, 1]$ and $Z_t \in \mathcal{Z}$
- Player observes Y_t and Z_t

We do not observe the counterfactual:

How would the environment have generated Y_t if the player picked a different A_t?

Suppose we also observe other variables that might help us learn this.

Causal Inference with Interventions via Multi-Armed Bandits

Multi-Armed Bandits with Post-Action Contexts

Fix a set of allowable *actions* $\mathcal{A}$ and *post-action contexts* $\mathcal{Z}$.

For each interaction $t \in [T]$:

- Player selects $A_t \in \mathcal{A}$
- Environment generates $Y_t \in [0, 1]$ and $Z_t \in \mathcal{Z}$
- Player observes Y_t and Z_t

We do not observe the counterfactual:

How would the environment have generated Y_t if the player picked a different A_t?

Suppose we also observe other variables that might help us learn this.

Big Assumption: We already know (learned offline) the marginal distribution of Z for each $a \in \mathcal{A}$.

Causal Inference with Interventions via Multi-Armed Bandits

Multi-Armed Bandits with Post-Action Contexts

Fix a set of allowable *actions* $\mathcal{A}$ and *post-action contexts* $\mathcal{Z}$.

For each interaction $t \in [T]$:

- Player selects $A_t \in \mathcal{A}$
- Environment generates $Y_t \in [0, 1]$ and $Z_t \in \mathcal{Z}$
- Player observes Y_t and Z_t

We do not observe the counterfactual:

How would the environment have generated Y_t if the player picked a different A_t?

Suppose we also observe other variables that might help us learn this.

Big Assumption: We already know (learned offline) the marginal distribution of Z for each $a \in \mathcal{A}$.

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

Measuring Performance – Formalized

Measuring Performance – Formalized

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

Measuring Performance – Formalized

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 (not counterfactuals) to actions.

Measuring Performance – Formalized

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 (not counterfactuals) to actions.

Together, they induce a distribution on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})^T$.

Measuring Performance – Formalized

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 (not counterfactuals) to actions.

Together, they induce a distribution on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})^T$.

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

Measuring Performance – Formalized

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 (not counterfactuals) to actions.

Together, they induce a distribution on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})^T$.

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

Important to note that...

Measuring Performance – Formalized

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 (not counterfactuals) to actions.

Together, they induce a distribution on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})^T$.

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

Important to note that...

We do not assume there is an "observational" distribution over $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$.

Measuring Performance – Formalized

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 (not counterfactuals) to actions.

Together, they induce a distribution on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})^T$.

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

Important to note that...

We do not assume there is an "observational" distribution over $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$.

We do not assume that ν satisfies any properties on any causal graph.

Measuring Performance – Formalized

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 (not counterfactuals) to actions.

Together, they induce a distribution on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})^T$.

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

Important to note that...

- We do not assume there is an "observational" distribution over $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$.
- We do not assume that ν satisfies any properties on any causal graph.
- We do not assume that $\mathcal{A}$ corresponds to "hard" interventions.

Measuring Performance – Formalized

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 (not counterfactuals) to actions.

Together, they induce a distribution on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})^T$.

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

Important to note that...

- We do not assume there is an "observational" distribution over $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$.
- We do not assume that ν satisfies any properties on any causal graph.
- We do not assume that $\mathcal{A}$ corresponds to "hard" interventions.

...but our framework includes these settings!

Measuring Performance – Formalized

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 (not counterfactuals) to actions.

Together, they induce a distribution on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})^T$.

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

Important to note that...

- We do not assume there is an "observational" distribution over $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$.
- We do not assume that ν satisfies any properties on any causal graph.
- We do not assume that $\mathcal{A}$ corresponds to "hard" interventions.

...but our framework includes these settings!

Measuring Performance – Formalized

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 (not counterfactuals) to actions.

Together, they induce a distribution on $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})^T$.

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

Important to note that...

- We do not assume there is an "observational" distribution over $(\mathcal{A} \times \mathcal{Z} \times \mathcal{Y})$.
- We do not assume that ν satisfies any properties on any causal graph.
- We do not assume that $\mathcal{A}$ corresponds to "hard" interventions.

...but our framework includes these settings!

What do we actually achieve?

What do we actually achieve?

The Punchline: High-Level Overview of Results

The Punchline: High-Level Overview of Results

Existing assumption: all *causal parents* of Y were observed.

The Punchline: High-Level Overview of Results

Existing assumption: all *causal parents* of Y were observed.

We introduce the *conditionally benign property*, which is strictly weaker than observing all parents.

The Punchline: High-Level Overview of Results

Existing assumption: all *causal parents* of Y were observed.

We introduce the *conditionally benign property*, which is strictly weaker than observing all parents.

Existing methods are designed to do well when all causal parents are observed.

The Punchline: High-Level Overview of Results

Existing assumption: all *causal parents* of Y were observed.

We introduce the *conditionally benign property*, which is strictly weaker than observing all parents.

Existing methods are designed to do well when all causal parents are observed.

We show they can incur *worst-case regret* (linear in T) when the assumption fails.

The Punchline: High-Level Overview of Results

Existing assumption: all *causal parents* of Y were observed.

We introduce the *conditionally benign property*, which is strictly weaker than observing all parents.

Existing methods are designed to do well when all causal parents are observed.

We show they can incur *worst-case regret* (linear in T) when the assumption fails.

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

The Punchline: High-Level Overview of Results

Existing assumption: all *causal parents* of Y were observed.

We introduce the *conditionally benign property*, which is strictly weaker than observing all parents.

Existing methods are designed to do well when all causal parents are observed.

We show they can incur *worst-case regret* (linear in T) when the assumption fails.

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

We show optimality is impossible: realizing the benefits necessarily sacrifices worst-case performance.

The Punchline: High-Level Overview of Results

Existing assumption: all *causal parents* of Y were observed.

We introduce the *conditionally benign property*, which is strictly weaker than observing all parents.

Existing methods are designed to do well when all causal parents are observed.

We show they can incur *worst-case regret* (linear in T) when the assumption fails.

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

We show optimality is impossible: realizing the benefits necessarily sacrifices worst-case performance.

We provide a new algorithm that:

The Punchline: High-Level Overview of Results

Existing assumption: all *causal parents* of Y were observed.

We introduce the *conditionally benign property*, which is strictly weaker than observing all parents.

Existing methods are designed to do well when all causal parents are observed.

We show they can incur *worst-case regret* (linear in T) when the assumption fails.

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

We show optimality is impossible: realizing the benefits necessarily sacrifices worst-case performance.

We provide a new algorithm that:

a) achieves optimal performance when the environment is conditionally benign;

The Punchline: High-Level Overview of Results

Existing assumption: all *causal parents* of Y were observed.

We introduce the *conditionally benign property*, which is strictly weaker than observing all parents.

Existing methods are designed to do well when all causal parents are observed.

We show they can incur *worst-case regret* (linear in T) when the assumption fails.

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

We show optimality is impossible: realizing the benefits necessarily sacrifices worst-case performance.

We provide a new algorithm that:

a) achieves optimal performance when the environment is conditionally benign;

b) always achieves sublinear regret;

The Punchline: High-Level Overview of Results

Existing assumption: all *causal parents* of Y were observed.

We introduce the *conditionally benign property*, which is strictly weaker than observing all parents.

Existing methods are designed to do well when all causal parents are observed.

We show they can incur *worst-case regret* (linear in T) when the assumption fails.

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

We show optimality is impossible: realizing the benefits necessarily sacrifices worst-case performance.

We provide a new algorithm that:

a) achieves optimal performance when the environment is conditionally benign;

b) always achieves sublinear regret;

c) incurs optimal $\sqrt{T}$ regret for some cases where existing algorithms incur linear regret.

The Punchline: High-Level Overview of Results

Existing assumption: all *causal parents* of Y were observed.

We introduce the *conditionally benign property*, which is strictly weaker than observing all parents.

Existing methods are designed to do well when all causal parents are observed.

We show they can incur *worst-case regret* (linear in T) when the assumption fails.

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

We show optimality is impossible: realizing the benefits necessarily sacrifices worst-case performance.

We provide a new algorithm that:

a) achieves optimal performance when the environment is conditionally benign;

b) always achieves sublinear regret;

c) incurs optimal $\sqrt{T}$ regret for some cases where existing algorithms incur linear regret.

What are we precisely trying to adapt to?

What are we precisely trying to adapt to?

Existing Results

Existing Results

Classically known (Auer et al. 2002) that the minimax regret is

$$R_{\nu,\pi}(T) = \Theta\left(\sqrt{|\mathcal{A}|\,T}\right).$$

Existing Results

Classically known (Auer et al. 2002) that the minimax regret is

$$R_{\nu,\pi}(T) = \Theta\left(\sqrt{|\mathcal{A}|\,T}\right).$$

This is achieved by, for example, the UCB algorithm.

Existing Results

Classically known (Auer et al. 2002) that the minimax regret is

$$R_{\nu,\pi}(T) = \Theta\left(\sqrt{|\mathcal{A}|\,T}\right).$$

This is achieved by, for example, the UCB algorithm.

Recently shown (Lu et al. 2020) that if Z is all causal parents of Y (under ν) and we know $(\nu_a(Z))_{a\in\mathcal{A}}$,

$$R_{\nu,\pi}(T) = \tilde{\Theta}\left(\sqrt{|\mathcal{Z}|\,T}\right).$$

Existing Results

Classically known (Auer et al. 2002) that the minimax regret is

$$R_{\nu,\pi}(T) = \Theta\left(\sqrt{|\mathcal{A}|\,T}\right).$$

This is achieved by, for example, the UCB algorithm.

Recently shown (Lu et al. 2020) that if Z is all causal parents of Y (under ν) and we know $(\nu_a(Z))_{a\in\mathcal{A}}$,

$$R_{\nu,\pi}(T) = \tilde{\Theta}\left(\sqrt{|\mathcal{Z}|\,T}\right).$$

They provide an algorithm that achieves this: C-UCB.

Existing Results

Classically known (Auer et al. 2002) that the minimax regret is

$$R_{\nu,\pi}(T) = \Theta\left(\sqrt{|\mathcal{A}|\, T}\right).$$

This is achieved by, for example, the UCB algorithm.

Recently shown (Lu et al. 2020) that if Z is all causal parents of Y (under ν) and we know $(\nu_a(Z))_{a \in \mathcal{A}}$,

$$R_{\nu,\pi}(T) = \tilde{\Theta}\left(\sqrt{|\mathcal{Z}|\, T}\right).$$

They provide an algorithm that achieves this: C-UCB.

Intuition

Existing Results

Classically known (Auer et al. 2002) that the minimax regret is

$$R_{\nu,\pi}(T) = \Theta\left(\sqrt{|\mathcal{A}|\,T}\right).$$

This is achieved by, for example, the UCB algorithm.

Recently shown (Lu et al. 2020) that if Z is all causal parents of Y (under ν) and we know $(\nu_a(Z))_{a\in\mathcal{A}}$,

$$R_{\nu,\pi}(T) = \tilde{\Theta}\left(\sqrt{|\mathcal{Z}|\,T}\right).$$

They provide an algorithm that achieves this: C-UCB.

Intuition

In the worst case, each $a \in \mathcal{A}$ must be sufficiently explored to learn $\nu_a(Y)$.

Existing Results

Classically known (Auer et al. 2002) that the minimax regret is

$$R_{\nu,\pi}(T) = \Theta\left(\sqrt{|\mathcal{A}|\, T}\right).$$

This is achieved by, for example, the UCB algorithm.

Recently shown (Lu et al. 2020) that if Z is all causal parents of Y (under ν) and we know $(\nu_a(Z))_{a\in\mathcal{A}}$,

$$R_{\nu,\pi}(T) = \tilde{\Theta}\left(\sqrt{|\mathcal{Z}|\, T}\right).$$

They provide an algorithm that achieves this: C-UCB.

Intuition

In the worst case, each $a \in \mathcal{A}$ must be sufficiently explored to learn $\nu_a(Y)$.

When Z is the causal parents, it can be used to "block" A.

Existing Results

Classically known (Auer et al. 2002) that the minimax regret is

$$R_{\nu,\pi}(T) = \Theta\left(\sqrt{|\mathcal{A}|\,T}\right).$$

This is achieved by, for example, the UCB algorithm.

Recently shown (Lu et al. 2020) that if Z is all causal parents of Y (under ν) and we know $(\nu_a(Z))_{a\in\mathcal{A}}$,

$$R_{\nu,\pi}(T) = \tilde{\Theta}\left(\sqrt{|\mathcal{Z}|\,T}\right).$$

They provide an algorithm that achieves this: C-UCB.

Intuition

In the worst case, each $a \in \mathcal{A}$ must be sufficiently explored to learn $\nu_a(Y)$.
When Z is the causal parents, it can be used to "block" A.
When $|\mathcal{Z}| \ll |\mathcal{A}|$, share information from Z to learn $\nu_a(Y)$ without selecting $A = a$.

Existing Results

Classically known (Auer et al. 2002) that the minimax regret is

$$R_{\nu,\pi}(T) = \Theta\left(\sqrt{|\mathcal{A}|\, T}\right).$$

This is achieved by, for example, the UCB algorithm.

Recently shown (Lu et al. 2020) that if Z is all causal parents of Y (under ν) and we know $(\nu_a(Z))_{a\in\mathcal{A}}$,

$$R_{\nu,\pi}(T) = \tilde{\Theta}\left(\sqrt{|\mathcal{Z}|\, T}\right).$$

They provide an algorithm that achieves this: C-UCB.

Intuition

In the worst case, each $a \in \mathcal{A}$ must be sufficiently explored to learn $\nu_a(Y)$.
When Z is the causal parents, it can be used to "block" A.
When $|\mathcal{Z}| \ll |\mathcal{A}|$, share information from Z to learn $\nu_a(Y)$ without selecting $A = a$.

Conditionally Benign Property

Conditionally Benign Property

When can we use Z to learn counterfactuals?

Conditionally Benign Property

When can we use Z to learn counterfactuals?

When the choice of A doesn't affect the outcome of Y conditional on Z.

Conditionally Benign Property

When can we use Z to learn counterfactuals?

When the choice of A doesn't affect the outcome of Y conditional on Z.

Definition (informal)

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

Conditionally Benign Property

When can we use Z to learn counterfactuals?

When the choice of A doesn't affect the outcome of Y conditional on Z.

Definition (informal)

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

Theorem (Refined Thm 1 of Lu et al. 2020)

For any $\mathcal{A}$, $\mathcal{Z}$, T, and conditionally benign ν, $R_{\nu,\text{C-UCB}}(T) \leq 2|\mathcal{Z}| + 6\sqrt{|\mathcal{Z}|T\log T} + (\log T)\sqrt{2T}$.

Conditionally Benign Property

When can we use Z to learn counterfactuals?

When the choice of A doesn't affect the outcome of Y conditional on Z.

Definition (informal)

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

Theorem (Refined Thm 1 of Lu et al. 2020)

For any $\mathcal{A}$, $\mathcal{Z}$, T, and conditionally benign ν, $R_{\nu,\text{C-UCB}}(T) \leq 2|\mathcal{Z}| + 6\sqrt{|\mathcal{Z}|T\log T} + (\log T)\sqrt{2T}$.

No causal structure or assumptions needed!

Conditionally Benign Property

When can we use Z to learn counterfactuals?

When the choice of A doesn't affect the outcome of Y conditional on Z.

Definition (informal)

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

Theorem (Refined Thm 1 of Lu et al. 2020)

For any $\mathcal{A}$, $\mathcal{Z}$, T, and conditionally benign ν, $R_{\nu,\text{C-UCB}}(T) \leq 2|\mathcal{Z}| + 6\sqrt{|\mathcal{Z}| T \log T} + (\log T)\sqrt{2T}$.

No causal structure or assumptions needed!

Connection to Causality I

Connection to Causality I

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

Connection to Causality I

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

(a)

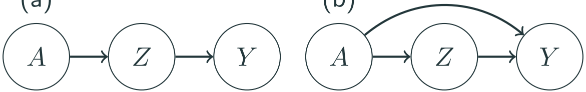
$$A \longrightarrow Z \longrightarrow Y$$

(a) conditionally benign and d-separated

Connection to Causality I

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

(a) (b)

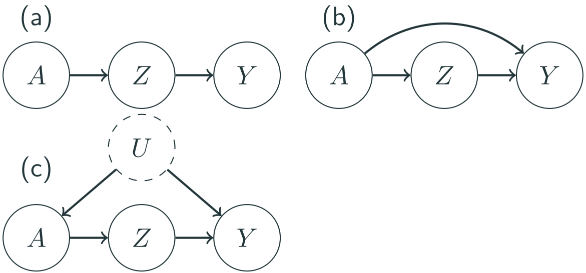


(a) conditionally benign and d-separated

(b) not conditionally benign

Connection to Causality I

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



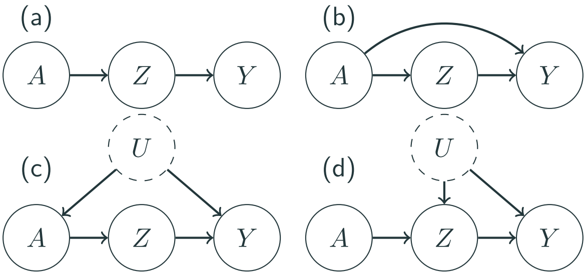
(a) conditionally benign and d-separated

(b) not conditionally benign

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

Connection to Causality I

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

Connection to Causality II

Connection to Causality II

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

Connection to Causality II

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 $\mathcal{A}$.

Connection to Causality II

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 $\mathcal{A}$.

Connection to Causality II

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 $\mathcal{A}_0$.

Connection to Causality II

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 $\mathcal{G}_{\overline{A}}$.

Why is this problem hard?

Why is this problem hard?

Adapting with Partial Feedback

Adapting with Partial Feedback

With full-information feedback, adaptivity is “easier”.

Adapting with Partial Feedback

With full-information feedback, adaptivity is "easier".

If I have access to policies $\pi_1, \ldots, \pi_N$, Bayesian posterior aggregation gives a π such that

$$R_{\nu,\pi}(T) \leq \min_{i \in [N]} R_{\nu,\pi_i}(T) + \sqrt{T \log N}.$$

Adapting with Partial Feedback

With full-information feedback, adaptivity is "easier".

If I have access to policies $\pi_1, \ldots, \pi_N$, Bayesian posterior aggregation gives a π such that

$$R_{\nu,\pi}(T) \leq \min_{i \in [N]} R_{\nu,\pi_i}(T) + \sqrt{T \log N}.$$

This works because π_i has the same regret regardless of what algorithm was played.

Adapting with Partial Feedback

With full-information feedback, adaptivity is "easier".

If I have access to policies $\pi_1, \ldots, \pi_N$, Bayesian posterior aggregation gives a π such that

$$R_{\nu,\pi}(T) \leq \min_{i \in [N]} R_{\nu,\pi_i}(T) + \sqrt{T \log N}.$$

This works because π_i has the same regret regardless of what algorithm was played.

This is no longer true with partial feedback!

Adapting with Partial Feedback

With full-information feedback, adaptivity is "easier".

If I have access to policies $\pi_1, \dots, \pi_N$, Bayesian posterior aggregation gives a π such that

$$R_{\nu,\pi}(T) \leq \min_{i \in [N]} R_{\nu,\pi_i}(T) + \sqrt{T \log N}.$$

This works because π_i has the same regret regardless of what algorithm was played.

This is no longer true with partial feedback!

Basic Questions

Adapting with Partial Feedback

With full-information feedback, adaptivity is "easier".

If I have access to policies $\pi_1, \dots, \pi_N$, Bayesian posterior aggregation gives a π such that

$$R_{\nu,\pi}(T) \leq \min_{i \in [N]} R_{\nu,\pi_i}(T) + \sqrt{T \log N}.$$

This works because π_i has the same regret regardless of what algorithm was played.

This is no longer true with partial feedback!

Basic Questions

Is there a generic way to aggregate bandit algorithms?

Adapting with Partial Feedback

With full-information feedback, adaptivity is "easier".

If I have access to policies $\pi_1, \ldots, \pi_N$, Bayesian posterior aggregation gives a π such that

$$R_{\nu,\pi}(T) \leq \min_{i \in [N]} R_{\nu,\pi_i}(T) + \sqrt{T \log N}.$$

This works because π_i has the same regret regardless of what algorithm was played.

This is no longer true with partial feedback!

Basic Questions

Is there a generic way to aggregate bandit algorithms?

What is the optimal adaptive performance possible for bandit algorithms?

Adapting with Partial Feedback

With full-information feedback, adaptivity is "easier".

If I have access to policies $\pi_1, \ldots, \pi_N$, Bayesian posterior aggregation gives a π such that

$$R_{\nu,\pi}(T) \leq \min_{i \in [N]} R_{\nu,\pi_i}(T) + \sqrt{T \log N}.$$

This works because π_i has the same regret regardless of what algorithm was played.

This is no longer true with partial feedback!

Basic Questions

Is there a generic way to aggregate bandit algorithms?

What is the optimal adaptive performance possible for bandit algorithms?

We provide partial, preliminary answers to these questions.

Adapting with Partial Feedback

With full-information feedback, adaptivity is “easier”.

If I have access to policies $\pi_1, \dots, \pi_N$, Bayesian posterior aggregation gives a π such that

$$R_{\nu,\pi}(T) \le \min_{i \in [N]} R_{\nu,\pi_i}(T) + \sqrt{T \log N}.$$

This works because π_i has the same regret regardless of what algorithm was played.

This is no longer true with partial feedback!

Basic Questions

Is there a generic way to aggregate bandit algorithms?

What is the optimal adaptive performance possible for bandit algorithms?

We provide partial, preliminary answers to these questions.

Lower Bounds on Adaptivity

Lower Bounds on Adaptivity

First, we show that existing algorithms do not adapt to failure of assumptions.

Lower Bounds on Adaptivity

First, we show that existing algorithms do not adapt to failure of assumptions.

Theorem

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

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

Lower Bounds on Adaptivity

First, we show that existing algorithms do not adapt to failure of assumptions.

Theorem

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

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

Next, we show that optimal adaptation to the conditionally benign property is impossible.

Lower Bounds on Adaptivity

First, we show that existing algorithms do not adapt to failure of assumptions.

Theorem

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

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

Next, we show that optimal adaptation to the conditionally benign property is impossible.
An algorithm $\mathfrak{a}$ maps T, $\mathcal{A}$, $\mathcal{Z}$, and $(\nu_a(Z))_{a\in\mathcal{A}}$ to a policy π.

Lower Bounds on Adaptivity

First, we show that existing algorithms do not adapt to failure of assumptions.

Theorem

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

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

Next, we show that optimal adaptation to the conditionally benign property is impossible.
An algorithm $\mathfrak{a}$ maps T, $\mathcal{A}$, $\mathcal{Z}$, and $(\nu_a(Z))_{a\in\mathcal{A}}$ to a policy π.

Theorem

If $\mathfrak{a}$ always satisfies $R_{\nu,\mathfrak{a}}(T) \leq \sqrt{|\mathcal{A}|T}$,
there exists a conditionally benign ν such that $R_{\nu,\mathfrak{a}}(T) \geq \sqrt{|\mathcal{A}|T}$.

Lower Bounds on Adaptivity

First, we show that existing algorithms do not adapt to failure of assumptions.

Theorem

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

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

Next, we show that optimal adaptation to the conditionally benign property is impossible.
An algorithm $\mathfrak{a}$ maps T, $\mathcal{A}$, $\mathcal{Z}$, and $(\nu_a(Z))_{a\in\mathcal{A}}$ to a policy π.

Theorem

If $\mathfrak{a}$ always satisfies $R_{\nu,\mathfrak{a}}(T) \leq \sqrt{|\mathcal{A}|T}$,
there exists a conditionally benign ν such that $R_{\nu,\mathfrak{a}}(T) \geq \sqrt{|\mathcal{A}|T}$.

Can we do any better?

Can we do any better?

Upper Bounds for HAC-UCB

We provide a new algorithm HAC-UCB that achieves non-trivial adaptivity.

Upper Bounds for HAC-UCB

We provide a new algorithm HAC-UCB that achieves non-trivial adaptivity.

Theorem

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

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

and if ν is conditionally benign,

$$R_{\nu,\text{HAC-UCB}}(T) \leq 2(|\mathcal{A}| + |\mathcal{Z}|) + 6\sqrt{|\mathcal{Z}|T\log T} + 2(\log T)\sqrt{T}\,.$$

Upper Bounds for HAC-UCB

We provide a new algorithm HAC-UCB that achieves non-trivial adaptivity.

Theorem

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

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

and if ν is conditionally benign,

$$R_{\nu,\text{HAC-UCB}}(T) \leq 2(|\mathcal{A}| + |\mathcal{Z}|) + 6\sqrt{|\mathcal{Z}|T\log T} + 2(\log T)\sqrt{T}\,.$$

Theorem

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

$$R_{\nu,\text{C-UCB}}(T) \geq \Omega(T) \quad \text{but} \quad R_{\nu,\text{HAC-UCB}}(T) \leq \tilde{O}(\sqrt{|\mathcal{A}|T}).$$

Upper Bounds for HAC-UCB

We provide a new algorithm HAC-UCB that achieves non-trivial adaptivity.

Theorem

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

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

and if ν is conditionally benign,

$$R_{\nu,\text{HAC-UCB}}(T) \leq 2(|\mathcal{A}| + |\mathcal{Z}|) + 6\sqrt{|\mathcal{Z}|T\log T} + 2(\log T)\sqrt{T}\,.$$

Theorem

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

$$R_{\nu,\text{C-UCB}}(T) \geq \Omega(T) \quad \text{but} \quad R_{\nu,\text{HAC-UCB}}(T) \leq \tilde{O}(\sqrt{|\mathcal{A}|T}).$$

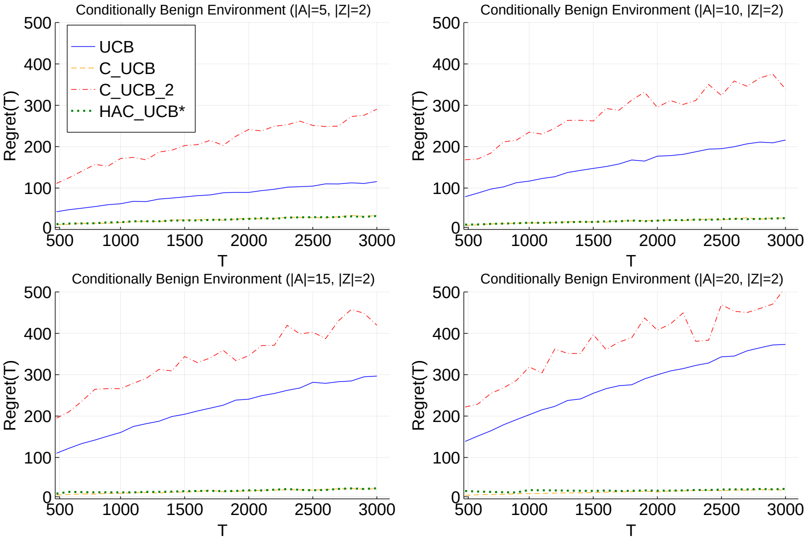
Empirical Investigation of Guarantees I

Empirical Investigation of Guarantees I

First, both C-UCB and HAC-UCB do better for conditionally benign ν.

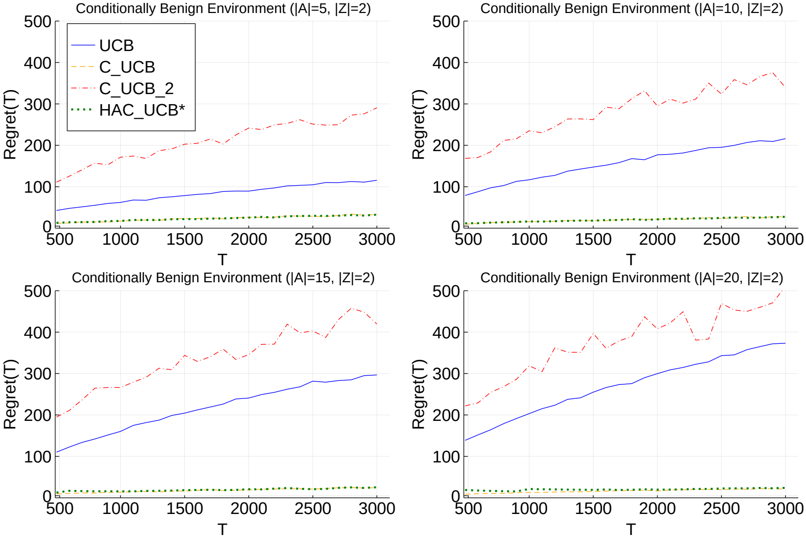
Empirical Investigation of Guarantees I

First, both C-UCB and HAC-UCB do better for conditionally benign ν.



Empirical Investigation of Guarantees I

First, both C-UCB and HAC-UCB do better for conditionally benign ν.

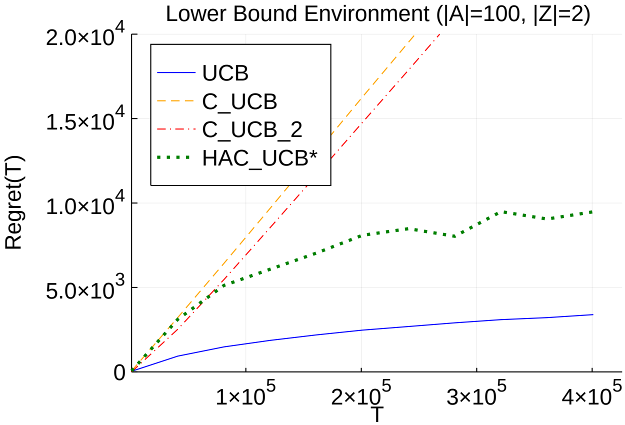


Empirical Investigation of Guarantees II

Second, C-UCB fails to do well for worst-case ν while HAC-UCB partially adapts.

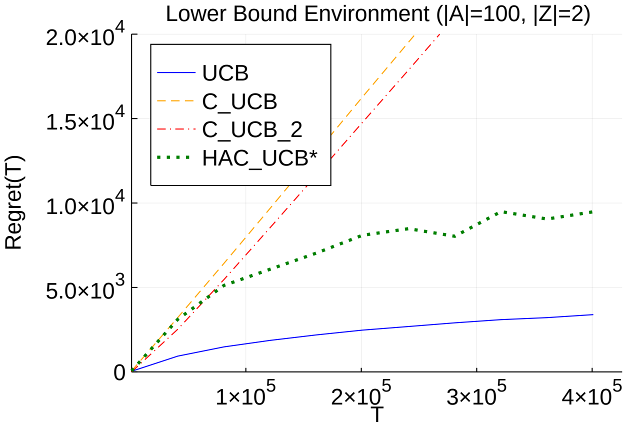
Empirical Investigation of Guarantees II

Second, C-UCB fails to do well for worst-case ν while HAC-UCB partially adapts.



Empirical Investigation of Guarantees II

Second, C-UCB fails to do well for worst-case ν while HAC-UCB partially adapts.



How do we achieve this?

How do we achieve this?

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 $\text{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(1/\delta)/N_t(a)}$

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 $\text{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(1/\delta)/N_t(a)}$
- 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 $\text{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(1/\delta)/N_t(a)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \text{UCB}_t(a)$

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 $\text{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(1/\delta)/N_t(a)}$
- 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 $\text{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(1/\delta)/N_t(a)}$
- 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 $\text{UCB}_t(z) = \hat{\mu}_t(z) + \sqrt{\log(1/\delta)/N_t(z)}$

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 $\mathrm{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(1/\delta)/N_t(a)}$
- 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 $\mathrm{UCB}_t(z) = \hat{\mu}_t(z) + \sqrt{\log(1/\delta)/N_t(z)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \sum_{z \in \mathcal{Z}} \mathrm{UCB}_t(z) \mathbb{P}_{\nu_a}[Z = z]$

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 $\mathrm{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(1/\delta)/N_t(a)}$
- 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 $\mathrm{UCB}_t(z) = \hat{\mu}_t(z) + \sqrt{\log(1/\delta)/N_t(z)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \sum_{z \in \mathcal{Z}} \mathrm{UCB}_t(z) \mathbb{P}_{\nu_a}[Z = z]$

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 $\mathsf{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(1/\delta)/N_t(a)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \mathsf{UCB}_t(a)$

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 $\mathsf{UCB}_t(z) = \hat{\mu}_t(z) + \sqrt{\log(1/\delta)/N_t(z)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \sum_{z \in \mathcal{Z}} \mathsf{UCB}_t(z)\mathbb{P}_{\nu_a}[Z = z]$

Why does this work?

If all parents are observed (more generally, ν is conditionally benign),

$$\sum_{z \in \mathcal{Z}} \mathsf{UCB}_t(z)\mathbb{P}_{\nu_a}[Z = z] \approx \mathsf{UCB}_t(a),$$

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 $\mathsf{UCB}_t(a) = \hat{\mu}_t(a) + \sqrt{\log(1/\delta)/N_t(a)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \mathsf{UCB}_t(a)$

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 $\mathsf{UCB}_t(z) = \hat{\mu}_t(z) + \sqrt{\log(1/\delta)/N_t(z)}$
- Play $A_t = \arg\max_{a \in \mathcal{A}} \sum_{z \in \mathcal{Z}} \mathsf{UCB}_t(z) \mathbb{P}_{\nu_a}[Z = z]$

Why does this work?

If all parents are observed (more generally, ν is conditionally benign),

$$\sum_{z \in \mathcal{Z}} \mathsf{UCB}_t(z) \mathbb{P}_{\nu_a}[Z = z] \approx \mathsf{UCB}_t(a),$$

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) Uniformly Explore

Adapting with Hypothesis Testing: HAC-UCB

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

(1) Uniformly Explore

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

Adapting with Hypothesis Testing: HAC-UCB

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

(1) Uniformly Explore

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.

Adapting with Hypothesis Testing: HAC-UCB

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

(1) Uniformly Explore

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.

Adapting with Hypothesis Testing: HAC-UCB

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

(1) Uniformly Explore

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.

(2) Optimistically Play C-UCB

Adapting with Hypothesis Testing: HAC-UCB

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

(1) Uniformly Explore

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.

(2) Optimistically Play C-UCB

Play C-UCB until the following fails:

$$-\sum_{z \in \mathcal{Z}} \sqrt{2\log(1/\delta)/N_t(z)}\mathbb{P}_{\nu_a}[Z = z] \leq \mathsf{UCB}_t(a) - \sum_{z \in \mathcal{Z}} \mathsf{UCB}_t(z)\mathbb{P}_{\nu_a}[Z = z] \leq \sqrt{2\log(1/\delta)/N_t(a)}.$$

Adapting with Hypothesis Testing: HAC-UCB

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

(1) Uniformly Explore

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.

(2) Optimistically Play C-UCB

Play C-UCB until the following fails:

$$-\sum_{z \in \mathcal{Z}} \sqrt{2\log(1/\delta)/N_t(z)}\mathbb{P}_{\nu_a}[Z = z] \leq \mathsf{UCB}_t(a) - \sum_{z \in \mathcal{Z}} \mathsf{UCB}_t(z)\mathbb{P}_{\nu_a}[Z = z] \leq \sqrt{2\log(1/\delta)/N_t(a)}.$$

(3) Pessimistically Play UCB

Once the hypothesis test fails, switch to UCB forever.

Adapting with Hypothesis Testing: HAC-UCB

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

(1) Uniformly Explore

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.

(2) Optimistically Play C-UCB

Play C-UCB until the following fails:

$$-\sum_{z \in \mathcal{Z}} \sqrt{2\log(1/\delta)/N_t(z)}\mathbb{P}_{\nu_a}[Z = z] \leq \mathrm{UCB}_t(a) - \sum_{z \in \mathcal{Z}} \mathrm{UCB}_t(z)\mathbb{P}_{\nu_a}[Z = z] \leq \sqrt{2\log(1/\delta)/N_t(a)}.$$

(3) Pessimistically Play UCB

Once the hypothesis test fails, switch to UCB forever.

Proof Sketch for HAC-UCB

Proof Sketch for HAC-UCB

Hypothesis test must be robust enough to reject bad environments quickly,
but sensitive enough to realize benefits of conditionally benign.

Proof Sketch for HAC-UCB

Hypothesis test must be robust enough to reject bad environments quickly,
but sensitive enough to realize benefits of conditionally benign.

Conditionally Benign Environments

Proof Sketch for HAC-UCB

Hypothesis test must be robust enough to reject bad environments quickly,
but sensitive enough to realize benefits of conditionally benign.

Conditionally Benign Environments

When ν is conditionally benign, $\mathrm{UCB}_t(a) \approx \sum_{z \in \mathcal{Z}} \mathrm{UCB}_t(z) \mathbb{P}_{\nu_a}[Z = z]$.

Proof Sketch for HAC-UCB

Hypothesis test must be robust enough to reject bad environments quickly,
but sensitive enough to realize benefits of conditionally benign.

Conditionally Benign Environments

When ν is conditionally benign, $\text{UCB}_t(a) \approx \sum_{z \in \mathcal{Z}} \text{UCB}_t(z) \mathbb{P}_{\nu_a}[Z = z]$.
Under empirical mean concentration, hypothesis test passes for all $t \in [T]$ simultaneously.

Proof Sketch for HAC-UCB

Hypothesis test must be robust enough to reject bad environments quickly,
but sensitive enough to realize benefits of conditionally benign.

Conditionally Benign Environments

When ν is conditionally benign, $\mathrm{UCB}_t(a) \approx \sum_{z \in \mathcal{Z}} \mathrm{UCB}_t(z) \mathbb{P}_{\nu_a}[Z = z]$.

Under empirical mean concentration, hypothesis test passes for all $t \in [T]$ simultaneously.

Worst-Case Environments

Proof Sketch for HAC-UCB

Hypothesis test must be robust enough to reject bad environments quickly,
but sensitive enough to realize benefits of conditionally benign.

Conditionally Benign Environments

When ν is conditionally benign, $\mathsf{UCB}_t(a) \approx \sum_{z \in \mathcal{Z}} \mathsf{UCB}_t(z)\mathbb{P}_{\nu_a}[Z = z]$.
Under empirical mean concentration, hypothesis test passes for all $t \in [T]$ simultaneously.

Worst-Case Environments

Remains to bound regret on rounds where still playing C-UCB.

Proof Sketch for HAC-UCB

Hypothesis test must be robust enough to reject bad environments quickly,
but sensitive enough to realize benefits of conditionally benign.

Conditionally Benign Environments

When ν is conditionally benign, $\mathsf{UCB}_t(a) \approx \sum_{z \in \mathcal{Z}} \mathsf{UCB}_t(z) \mathbb{P}_{\nu_a}[Z = z]$.
Under empirical mean concentration, hypothesis test passes for all $t \in [T]$ simultaneously.

Worst-Case Environments

Remains to bound regret on rounds where still playing C-UCB.

$$\mathbb{E}_{\nu_{a^*}}[Y] - \mathbb{E}_{\nu_{A_t}}[Y] = \underbrace{\mathbb{E}_{\nu_{a^*}}[Y] - \sum_{z \in \mathcal{Z}} \mathsf{UCB}_t(z) \mathbb{P}_{\nu_{A_t}}[Z = z]} + \underbrace{\sum_{z \in \mathcal{Z}} \mathsf{UCB}_t(z) \mathbb{P}_{\nu_{A_t}}[Z = z] - \mathbb{E}_{\nu_{A_t}}[Y]}$$

Proof Sketch for HAC-UCB

Hypothesis test must be robust enough to reject bad environments quickly,
but sensitive enough to realize benefits of conditionally benign.

Conditionally Benign Environments

When ν is conditionally benign, $\mathsf{UCB}_t(a) \approx \sum_{z \in \mathcal{Z}} \mathsf{UCB}_t(z)\mathbb{P}_{\nu_a}[Z = z]$.
Under empirical mean concentration, hypothesis test passes for all $t \in [T]$ simultaneously.

Worst-Case Environments

Remains to bound regret on rounds where still playing C-UCB.

$$\mathbb{E}_{\nu_{a^*}}[Y] - \mathbb{E}_{\nu_{A_t}}[Y] = \underbrace{\mathbb{E}_{\nu_{a^*}}[Y] - \sum_{z \in \mathcal{Z}} \mathsf{UCB}_t(z)\mathbb{P}_{\nu_{A_t}}[Z = z]}_{\text{Concentration } + A_t \text{ plays C-UCB}} + \underbrace{\sum_{z \in \mathcal{Z}} \mathsf{UCB}_t(z)\mathbb{P}_{\nu_{A_t}}[Z = z] - \mathbb{E}_{\nu_{A_t}}[Y]}$$

Proof Sketch for HAC-UCB

Hypothesis test must be robust enough to reject bad environments quickly,
but sensitive enough to realize benefits of conditionally benign.

Conditionally Benign Environments

When ν is conditionally benign, $\mathrm{UCB}_t(a) \approx \sum_{z \in \mathcal{Z}} \mathrm{UCB}_t(z) \mathbb{P}_{\nu_a}[Z = z]$.

Under empirical mean concentration, hypothesis test passes for all $t \in [T]$ simultaneously.

Worst-Case Environments

Remains to bound regret on rounds where still playing C-UCB.

$$\mathbb{E}_{\nu_{a^*}}[Y] - \mathbb{E}_{\nu_{A_t}}[Y] = \underbrace{\mathbb{E}_{\nu_{a^*}}[Y] - \sum_{z \in \mathcal{Z}} \mathrm{UCB}_t(z) \mathbb{P}_{\nu_{A_t}}[Z = z]}_{\text{Concentration } + A_t \text{ plays C-UCB}} + \underbrace{\sum_{z \in \mathcal{Z}} \mathrm{UCB}_t(z) \mathbb{P}_{\nu_{A_t}}[Z = z] - \mathbb{E}_{\nu_{A_t}}[Y]}_{\text{Martingale argument}}$$

What does this mean for optimal adaptivity?

What does this mean for optimal adaptivity?

A Pareto Frontier of Adaptivity

A Pareto Frontier of Adaptivity

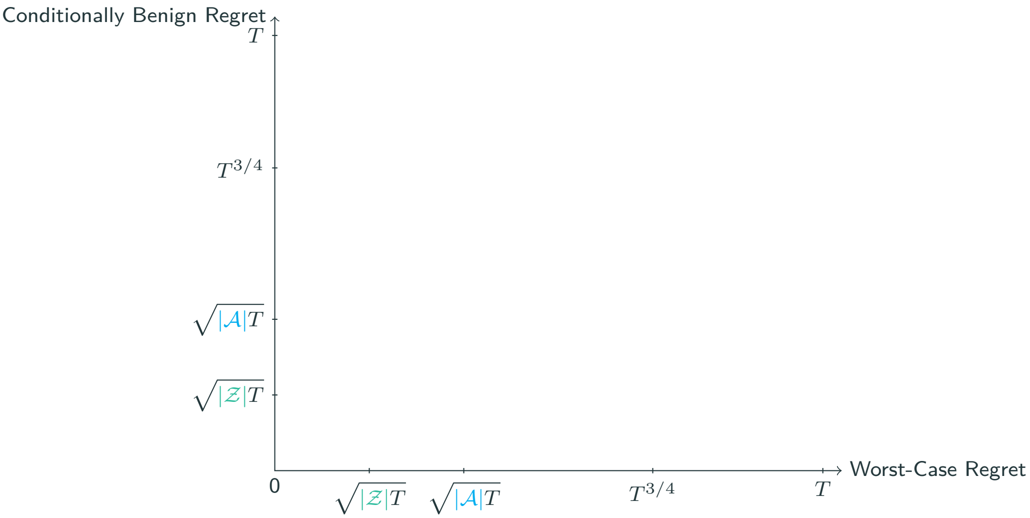
We know that optimal adaptivity is impossible.

A Pareto Frontier of Adaptivity

We know that optimal adaptivity is impossible. What else can we hope for?

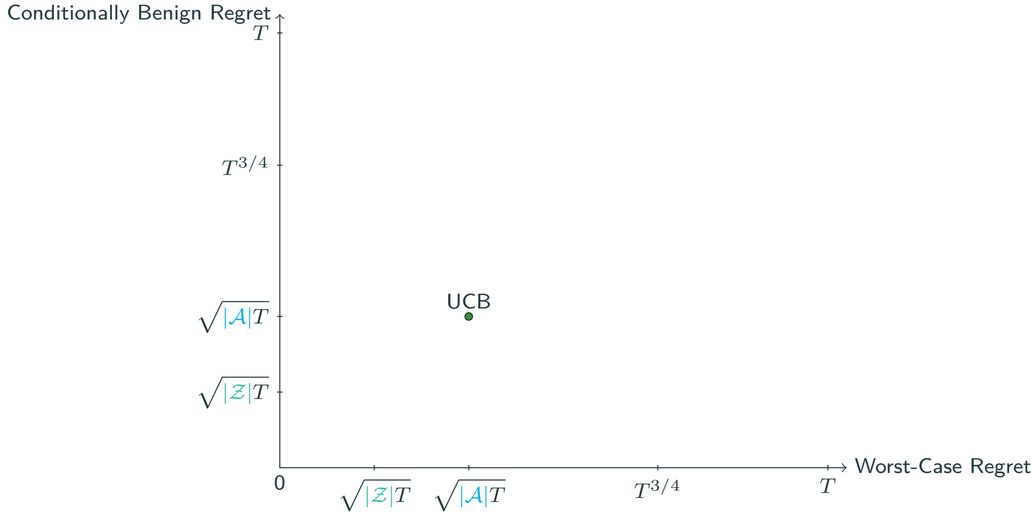
A Pareto Frontier of Adaptivity

We know that optimal adaptivity is impossible. What else can we hope for?



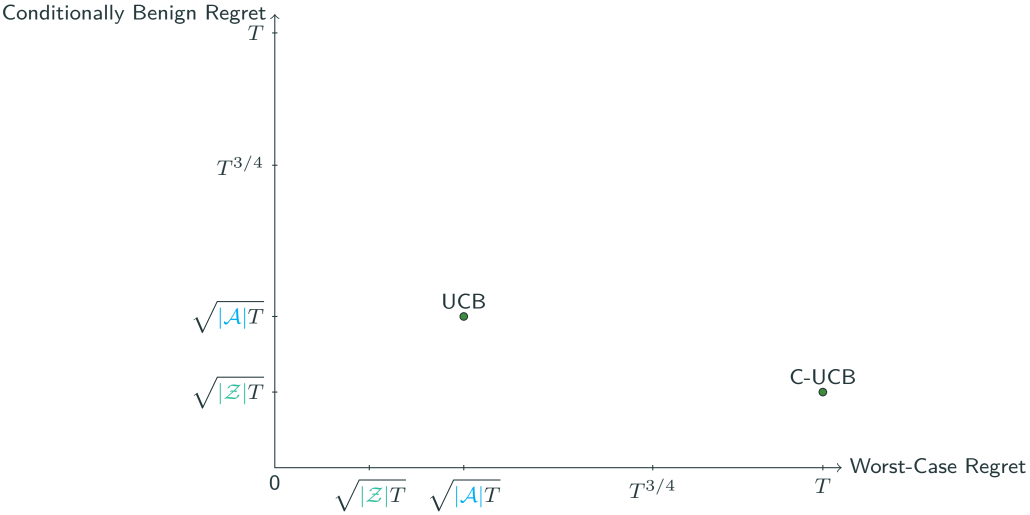
A Pareto Frontier of Adaptivity

We know that optimal adaptivity is impossible. What else can we hope for?



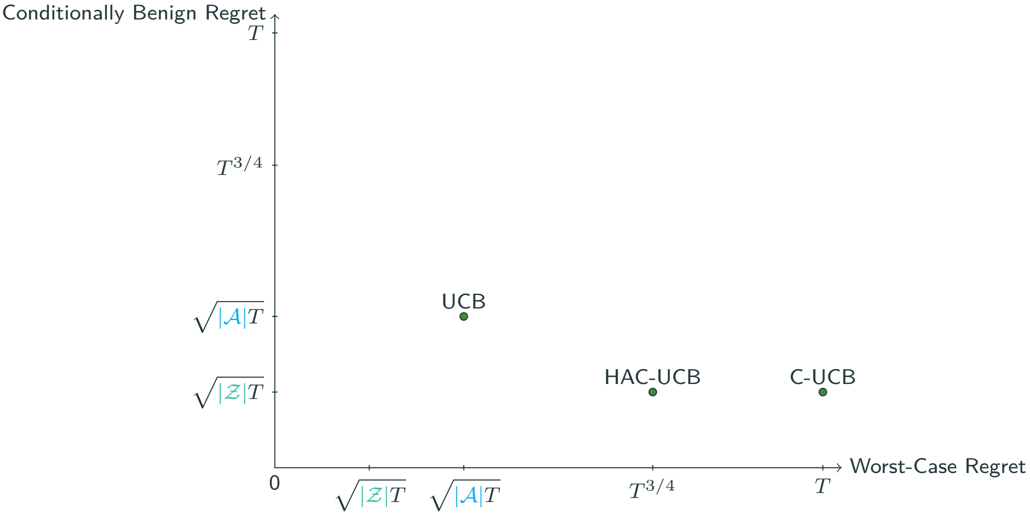
A Pareto Frontier of Adaptivity

We know that optimal adaptivity is impossible. What else can we hope for?



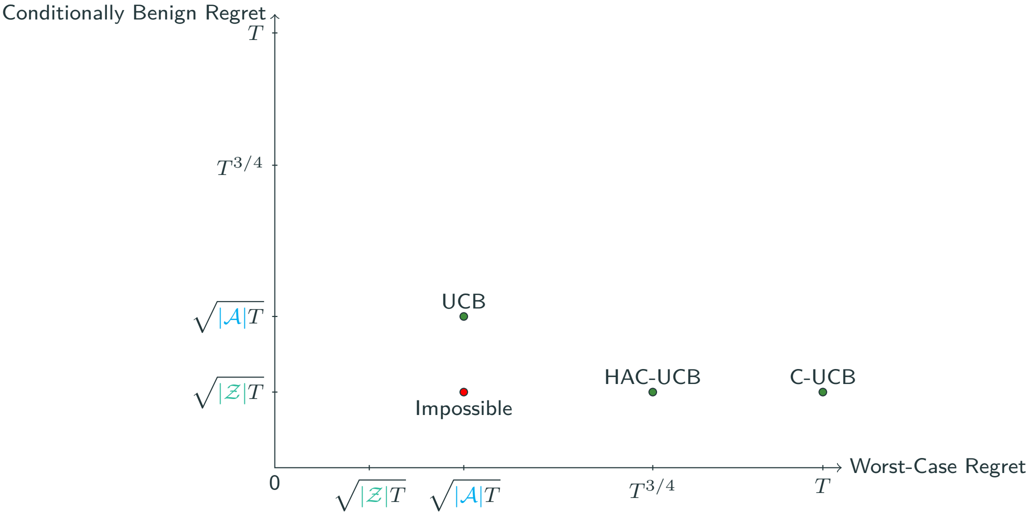
A Pareto Frontier of Adaptivity

We know that optimal adaptivity is impossible. What else can we hope for?



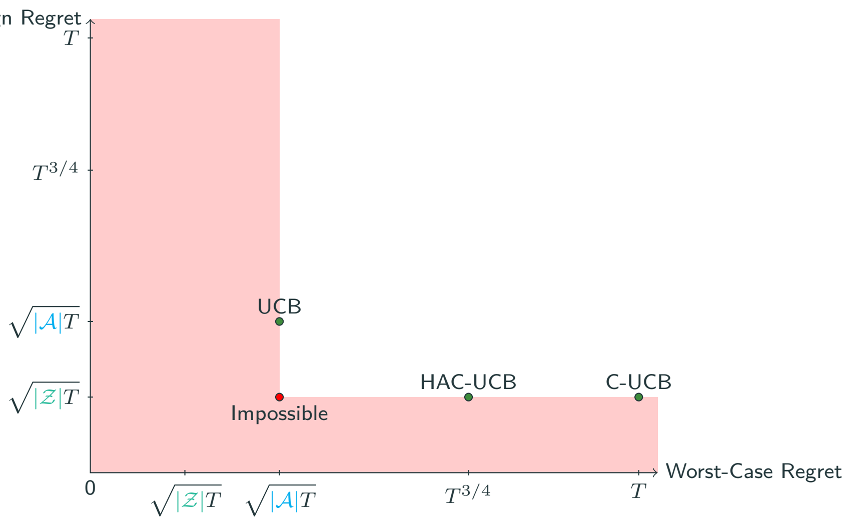
A Pareto Frontier of Adaptivity

We know that optimal adaptivity is impossible. What else can we hope for?



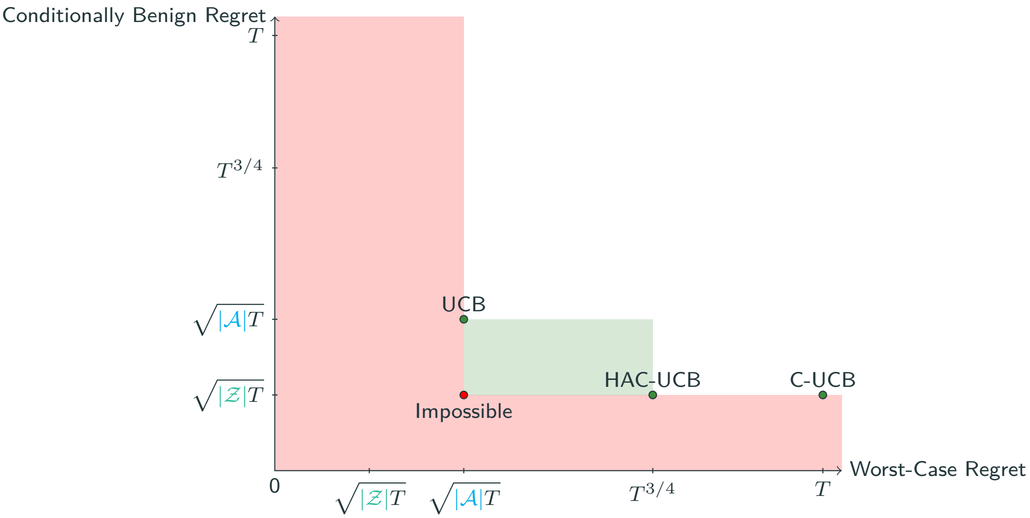
A Pareto Frontier of Adaptivity

We know that optimal adaptivity is impossible. What else can we hope for?



A Pareto Frontier of Adaptivity

We know that optimal adaptivity is impossible. What else can we hope for?



Proof Sketch for Lower Bounds

Proof Sketch for Lower Bounds

C-UCB Lower Bound

Proof Sketch for Lower Bounds

C-UCB Lower Bound

- If ν is not conditionally benign, the naive estimate of $\mathbb{E}[Y \mid Z]$ is biased.

Proof Sketch for Lower Bounds

C-UCB Lower Bound

- If ν is not conditionally benign, the naive estimate of $\mathbb{E}[Y \mid Z]$ is biased.
- In the worst-case, C-UCB only selects the optimal intervention finitely many times.

Proof Sketch for Lower Bounds

C-UCB Lower Bound

- If ν is not conditionally benign, the naive estimate of $\mathbb{E}[Y \mid Z]$ is biased.
- In the worst-case, C-UCB only selects the optimal intervention finitely many times.

An optimal algorithm must identify when the environment is not conditionally benign...

Proof Sketch for Lower Bounds

C-UCB Lower Bound

- If ν is not conditionally benign, the naive estimate of $\mathbb{E}[Y \mid Z]$ is biased.
- In the worst-case, C-UCB only selects the optimal intervention finitely many times.

An optimal algorithm must identify when the environment is not conditionally benign...
...but it may be possible to ignore "weakly" conditionally benign environments.

Proof Sketch for Lower Bounds

C-UCB Lower Bound

- If ν is not conditionally benign, the naive estimate of $\mathbb{E}[Y \mid Z]$ is biased.
- In the worst-case, C-UCB only selects the optimal intervention finitely many times.

An optimal algorithm must identify when the environment is not conditionally benign...

...but it may be possible to ignore "weakly" conditionally benign environments.

Minimax Lower Bound

Proof Sketch for Lower Bounds

C-UCB Lower Bound

- If ν is not conditionally benign, the naive estimate of $\mathbb{E}[Y \mid Z]$ is biased.
- In the worst-case, C-UCB only selects the optimal intervention finitely many times.

An optimal algorithm must identify when the environment is not conditionally benign...
...but it may be possible to ignore "weakly" conditionally benign environments.

Minimax Lower Bound

- It is possible to construct a conditionally benign environment that is indistinguishable from $|\mathcal{A}| - 1$ non-conditionally benign environments.

Proof Sketch for Lower Bounds

C-UCB Lower Bound

- If ν is not conditionally benign, the naive estimate of $\mathbb{E}[Y \mid Z]$ is biased.
- In the worst-case, C-UCB only selects the optimal intervention finitely many times.

An optimal algorithm must identify when the environment is not conditionally benign...
...but it may be possible to ignore "weakly" conditionally benign environments.

Minimax Lower Bound

- It is possible to construct a conditionally benign environment that is indistinguishable from $|\mathcal{A}| - 1$ non-conditionally benign environments.
- Any worst-case optimal algorithm must explore all interventions sufficiently often.

Proof Sketch for Lower Bounds

C-UCB Lower Bound

- If ν is not conditionally benign, the naive estimate of $\mathbb{E}[Y \mid Z]$ is biased.
- In the worst-case, C-UCB only selects the optimal intervention finitely many times.

An optimal algorithm must identify when the environment is not conditionally benign...
...but it may be possible to ignore "weakly" conditionally benign environments.

Minimax Lower Bound

- It is possible to construct a conditionally benign environment that is indistinguishable from $|\mathcal{A}| - 1$ non-conditionally benign environments.
- Any worst-case optimal algorithm must explore all interventions sufficiently often.

No algorithm can bypass this...

Proof Sketch for Lower Bounds

C-UCB Lower Bound

- If ν is not conditionally benign, the naive estimate of $\mathbb{E}[Y \mid Z]$ is biased.
- In the worst-case, C-UCB only selects the optimal intervention finitely many times.

An optimal algorithm must identify when the environment is not conditionally benign...

...but it may be possible to ignore "weakly" conditionally benign environments.

Minimax Lower Bound

- It is possible to construct a conditionally benign environment that is indistinguishable from $|\mathcal{A}| - 1$ non-conditionally benign environments.
- Any worst-case optimal algorithm must explore all interventions sufficiently often.

No algorithm can bypass this...

...but it may be possible to bypass it up to log factors.

Proof Sketch for Lower Bounds

C-UCB Lower Bound

- If ν is not conditionally benign, the naive estimate of $\mathbb{E}[Y \mid Z]$ is biased.
- In the worst-case, C-UCB only selects the optimal intervention finitely many times.

An optimal algorithm must identify when the environment is not conditionally benign...

...but it may be possible to ignore "weakly" conditionally benign environments.

Minimax Lower Bound

- It is possible to construct a conditionally benign environment that is indistinguishable from $|\mathcal{A}| - 1$ non-conditionally benign environments.
- Any worst-case optimal algorithm must explore all interventions sufficiently often.

No algorithm can bypass this...

...but it may be possible to bypass it up to log factors.

Conclusions

Conclusions

- We want to be agnostic to assumptions about how the data is generated.

Conclusions

- We want to be agnostic to assumptions about how the data is generated.
- Must intervene to achieve this for identifying the best intervention without confounding assumptions.

Conclusions

- We want to be agnostic to assumptions about how the data is generated.
- Must intervene to achieve this for identifying the best intervention without confounding assumptions.
- We formalize and generalize what assumptions is “benign” for this task.

Conclusions

- We want to be agnostic to assumptions about how the data is generated.
- Must intervene to achieve this for identifying the best intervention without confounding assumptions.
- We formalize and generalize what assumptions is "benign" for this task.
- We show existing algorithms fail strongly when assumptions fail.

Conclusions

- We want to be agnostic to assumptions about how the data is generated.
- Must intervene to achieve this for identifying the best intervention without confounding assumptions.
- We formalize and generalize what assumptions is "benign" for this task.
- We show existing algorithms fail strongly when assumptions fail.
- We show optimal adaptivity is impossible and formalize the Pareto trade-off for conditionally benign.

Conclusions

- We want to be agnostic to assumptions about how the data is generated.
- Must intervene to achieve this for identifying the best intervention without confounding assumptions.
- We formalize and generalize what assumptions is "benign" for this task.
- We show existing algorithms fail strongly when assumptions fail.
- We show optimal adaptivity is impossible and formalize the Pareto trade-off for conditionally benign.
- We provide HAC-UCB, which has non-trivial adaptivity.

Conclusions

- We want to be agnostic to assumptions about how the data is generated.
- Must intervene to achieve this for identifying the best intervention without confounding assumptions.
- We formalize and generalize what assumptions is "benign" for this task.
- We show existing algorithms fail strongly when assumptions fail.
- We show optimal adaptivity is impossible and formalize the Pareto trade-off for conditionally benign.
- We provide HAC-UCB, which has non-trivial adaptivity.
- Many wide-open and interesting problems to study in bandit adaptivity!

Conclusions

- We want to be agnostic to assumptions about how the data is generated.
- Must intervene to achieve this for identifying the best intervention without confounding assumptions.
- We formalize and generalize what assumptions is “benign” for this task.
- We show existing algorithms fail strongly when assumptions fail.
- We show optimal adaptivity is impossible and formalize the Pareto trade-off for conditionally benign.
- We provide HAC-UCB, which has non-trivial adaptivity.
- Many wide-open and interesting problems to study in bandit adaptivity!