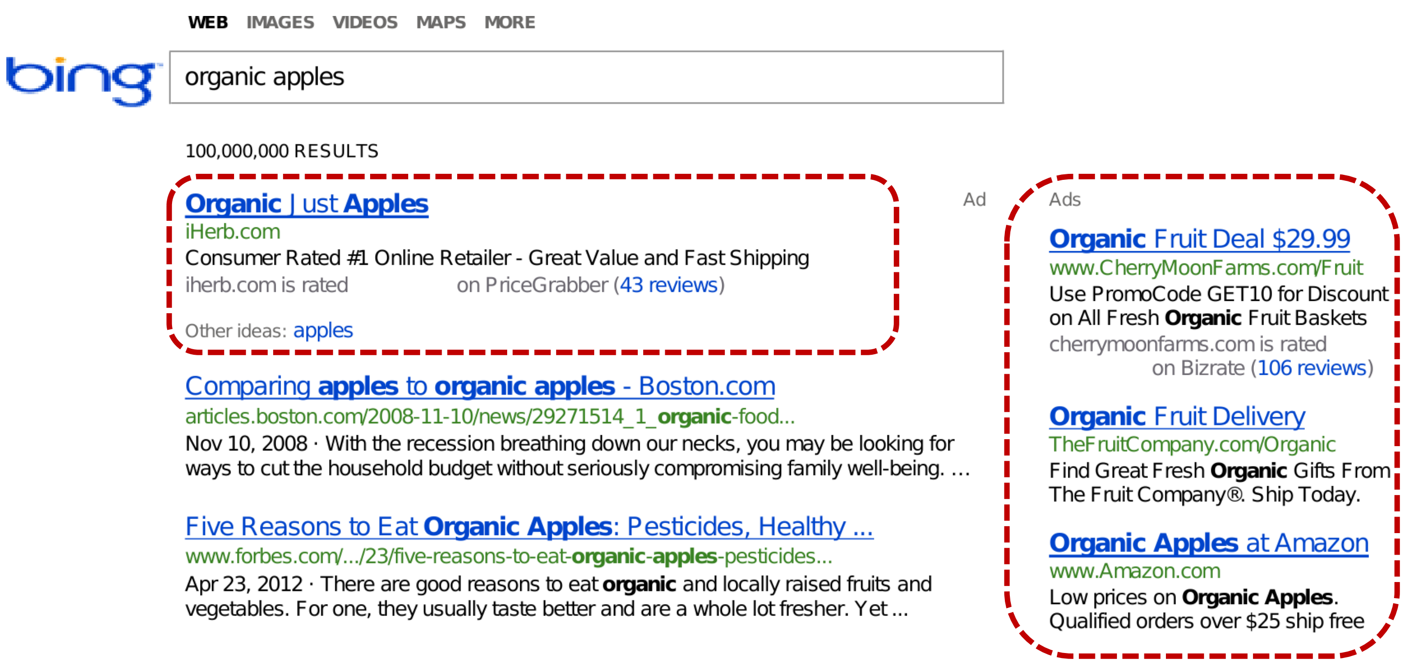
Causal Reasoning and Learning Systems

Elon Portugaly

MSR Cambridge UK / bing.ads

The pesky little ads



Work by

Léon Bottou

Jointly with:

Jonas Peters

Joaquin Quiñonero Candela

Denis Xavier Charles

D. Max Chickering

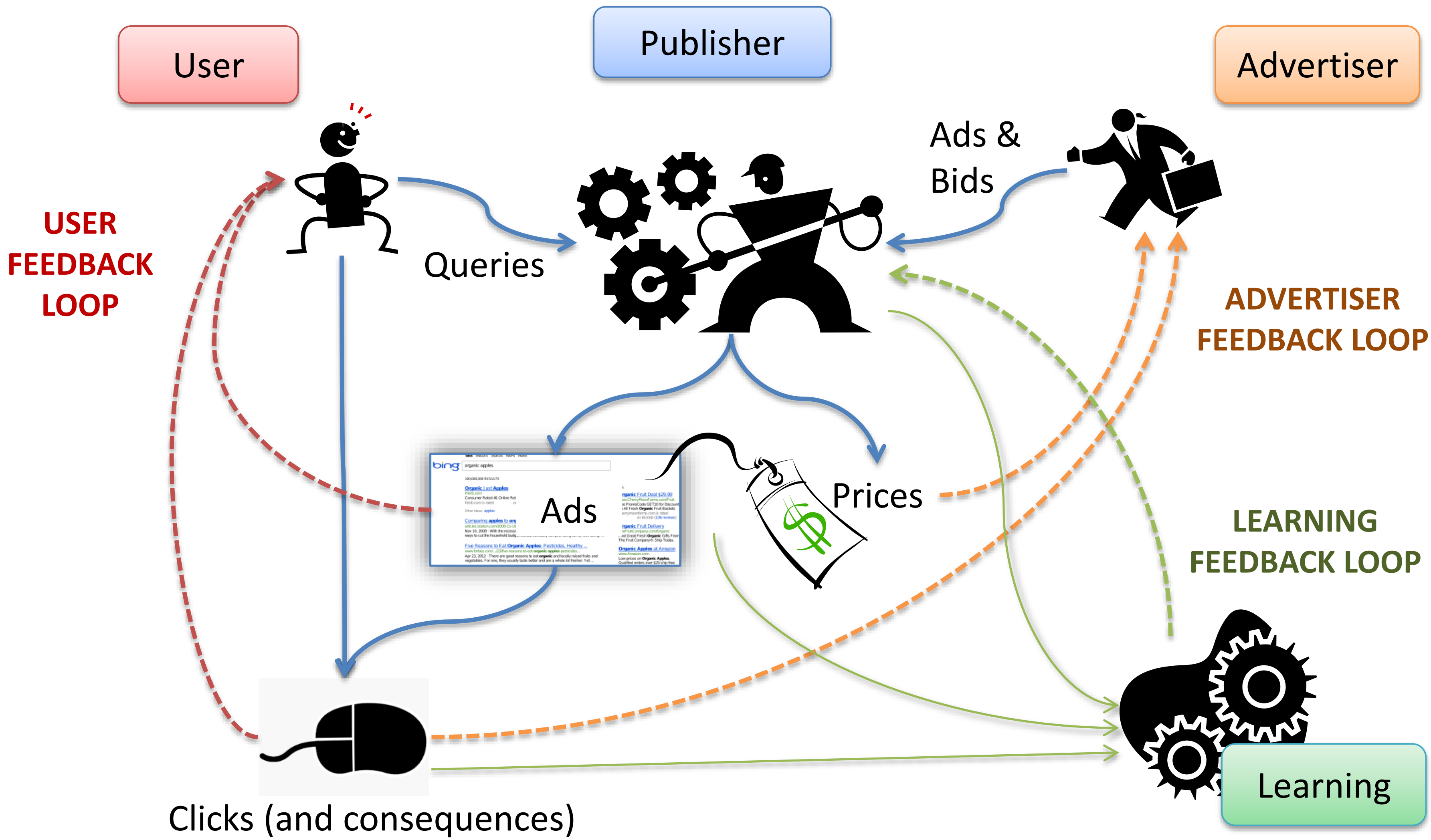
Elon Portugaly

Dipankar Ray

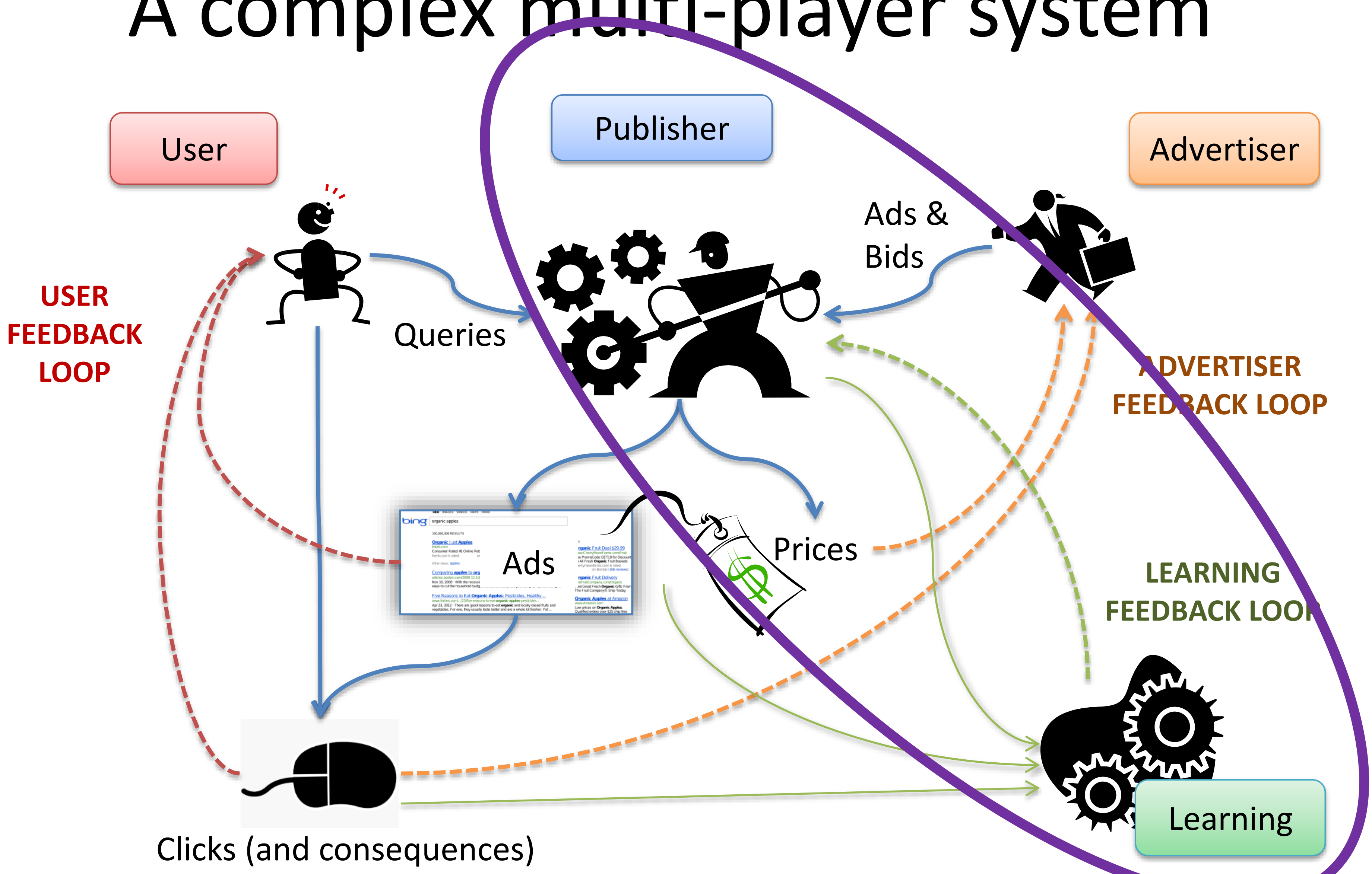
Patrice Simard

Ed Snelson

A complex multi-player system

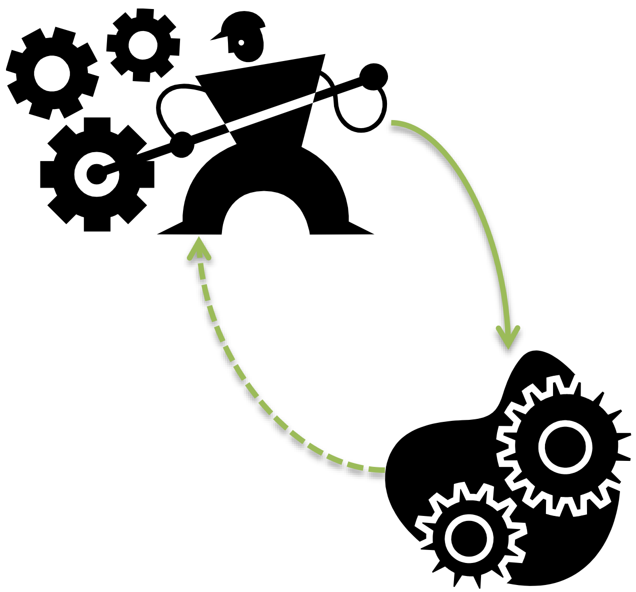


A complex multi-player system



Learning to run a marketplace

- Goal: improve marketplace machinery such that its long term revenue is maximal
- Approximate goal by improving multiple performance measures related to all players
- The learning machine
  is not a machine but
  is an organization with lots
  of people doing stuff and
  making decisions
  working in the dark

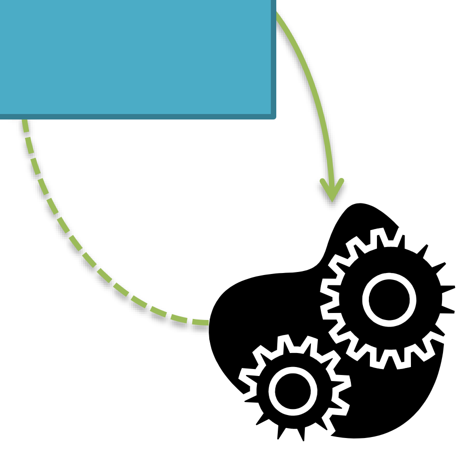


How can we help?

Learning to run a marketplace

- Goal: improve marketplace machinery such that its long term revenue is maximal
- Approximate goal by improving multiple performance measures related to all players
- Th
  is
  is an organization with lots of people doing stuff and making decisions working in the dark

> - Provide data for decision making
> - Automatically optimize parts of the system



How can we help?

The feedback loop problem (why exploration is necessary)

- **Shifting distributions**
  - Data is collected when the system operates in a certain way.
    The observed data follows a first distribution.
  - Collected data is used to justify actions that change the operating point.
    Newly observed data then follows a second distribution.
  - **Correlations observed on data following the first distribution**
    do not necessarily exist in the second distribution.
- **Often lead to vicious circles..**

Toy example

- **True conditional click probabilities**

	A1 (cheap jewelry)	**A2** (cheap autos)	**A3** (engagement rings)
Q1 (cheap diamonds)	7%	2%	9%
Q2 (news)	2%	2%	2%

Step 1: pick ads randomly.

$$Clicks = \frac{1}{2}\left(\frac{7\% + 2\% + 9\%}{3} + \frac{2\% + 2\% + 2\%}{3}\right) = 4\%$$

Toy example

- **Step 2: estimate click probabilities**
  - Build a model based on a single Boolean feature:
    F1 : "*query and ad have at least one word in common*"

	A1 (cheap jewelry)	**A2** (cheap autos)	**A3** (engagement rings)
Q1 (cheap diamonds)	7%	2%	9%
Q2 (news)	2%	2%	2%

$$P(Click|F1) = \frac{7\% + 2\%}{2} = 4.5\%$$

$$P(Click|\neg F1) = \frac{9\% + 2\% + 2\% + 2\%}{4} = 3.75\%$$

Toy example

- **Step 3: place ads according to estimated pclick.**

Q1: show A1 or A2. (predicted pclick 4.5% > 3.75%)

Q2: show A1, A2, or A3. (predicted pclick 3.75%)

	A1 (cheap jewelry)	**A2** (cheap autos)	**A3** (engagement rings)
Q1 (cheap diamonds)	7%	2%	~~9%~~
Q2 (news)	2%	2%	2%

$$Clicks = \frac{1}{2}\left(\frac{7\% + 2\%}{2} + \frac{2\% + 2\% + 2\%}{3}\right) = 3.25\%$$

Toy example

- **Step 4: re-estimate click probabilities with new data.**

	A1 (cheap jewelry)	**A2** (cheap autos)	**A3** (engagement rings)
Q1 (cheap diamonds)	7%	2%	~~9%~~
Q2 (news)	2%	2%	2%

$$P(Click|F1) = \frac{7\% + 2\%}{2} = 4.5\%$$

$$P(Click|\neg F1) = \frac{2\% + 2\% + 2\%}{3} = 2\%$$

- We keep selecting the same inferior ads. 
- Estimated click probabilities now seem more accurate. 

What is going wrong?

- Estimating Pclick using click data collected by showing random ads.

1 Feature F1 identifies relevant ads using a narrow criterion.

2 Feature F1 misses a very good ad for query Q1.

	A1	A2	A3
Q1	7%	2%	9%
Q2	2%	2%	2%

3 $P(\text{Click} \mid \neg F1)$ is pulled down by queries that do not click.

4 Ads for query Q1 are ranked incorrectly.

- Adding a feature F2 that singles out the case (Q1,A3)
  - would improve the pclick estimation metric.
  - would rank Q1 ads more adequately.

What is going wrong?

- Re-estimating Pclick using click data collected by showing ads suggested by the previous Pclick model.

	A1	A2	A3
Q1	7%	2%	~~9%~~
Q2	2%	2%	2%

In this data, A3 is never shown for query Q1.

$P(Click|\neg F1)$ seems more accurate because we have removed the case (Q1,A3)

- Adding a feature F2
  - would not improve the Pclick estimation on this data.
  - would not help ranking (Q1,A3) higher.
- Further feature engineering based on this data
  - would always result in eliminating more options, e.g. (Q1,A2).
  - would never result in recovering lost options, e.g. (Q1,A3).

We have created a black hole!

- **(Q,A) can be occasionally sucked by the black hole.**
  - All kinds of events can cause ads to disappear.
  - Sometimes, advertisers spend extra money to displace competitors.
- **(Q,A) can be born in the black hole.**
  - Ads newly entered by advertisers
  - Ads newly selected as eligible because of algorithmic improvements.

- **Exploration**
  - We should sometimes show ads that we would not normally show in order to train the click prediction model.

Counterfactuals, interventions, and randomization

Counterfactuals: Measuring something that did not happen

"How would the system have performed if, when the data was collected, we had intervened and used $P^*(C)$ instead of $P(C)$?"

In a randomized system, counterfactual estimation is possible

Interventions are a change in a distribution

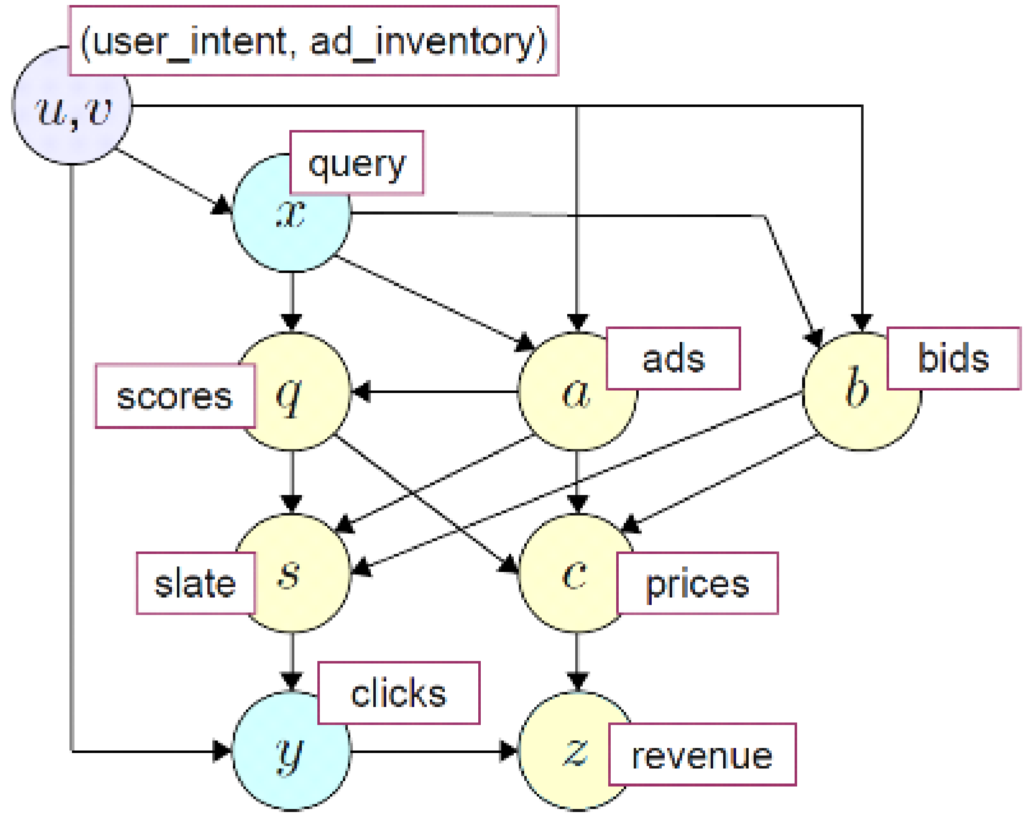
Estimating one distribution using data generated by another distribution

Learning procedure

- Collect data that describes the operation of the system during a past time period.
- Find changes that would have increased the performance of the system if they had been applied during the data collection period.
- Implement and verify…

Markov factorization

$$
\begin{aligned}
\mathbf{P}(\omega) = \mathbf{P}(u,v) \\
&\times \mathbf{P}(x \mid u) \\
&\times \mathbf{P}(a \mid x, v) \\
&\times \mathbf{P}(b \mid x, v) \\
&\times \mathbf{P}(q \mid x, a) \\
&\times \mathbf{P}(s \mid a, q, b) \\
&\times \mathbf{P}(c \mid a, q, b) \\
&\times \mathbf{P}(y \mid s, u) \\
&\times \mathbf{P}(z \mid y, c)
\end{aligned}
$$



Some variables are observed, some are not
Some factors are known, some are not
Some factors can be manipulated some can't

Markov interventions

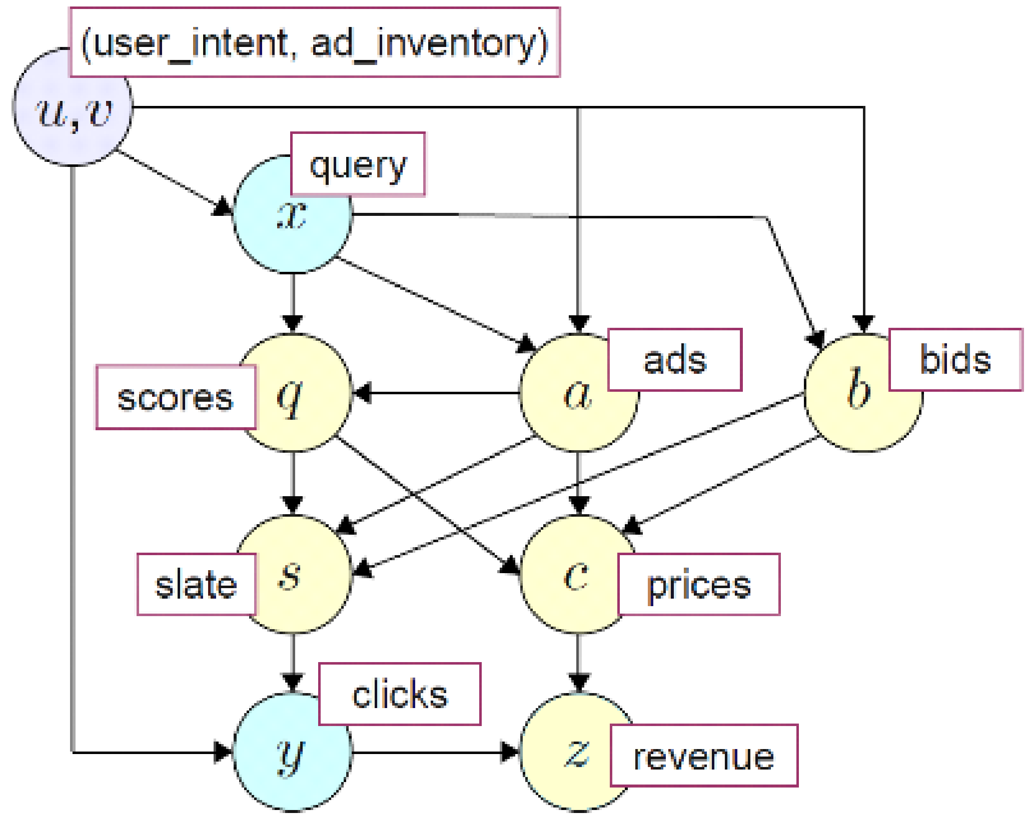
$$\begin{aligned} P^{*}(\omega) = \; & P(u,v) \\ & \times P(x \mid u) \\ & \times P(a \mid x,v) \\ & \times P(b \mid x,v) \\ & \times \cancel{P(q \mid x,a)} \quad P^{*}(q \mid x,a) \\ & \times P(s \mid a,q,b) \\ & \times P(c \mid a,q,b) \\ & \times P(y \mid s,u) \\ & \times P(z \mid y,c) \end{aligned}$$

Distribution under intervention

Many interrelated Bayes networks are born (Pearl, 2000)

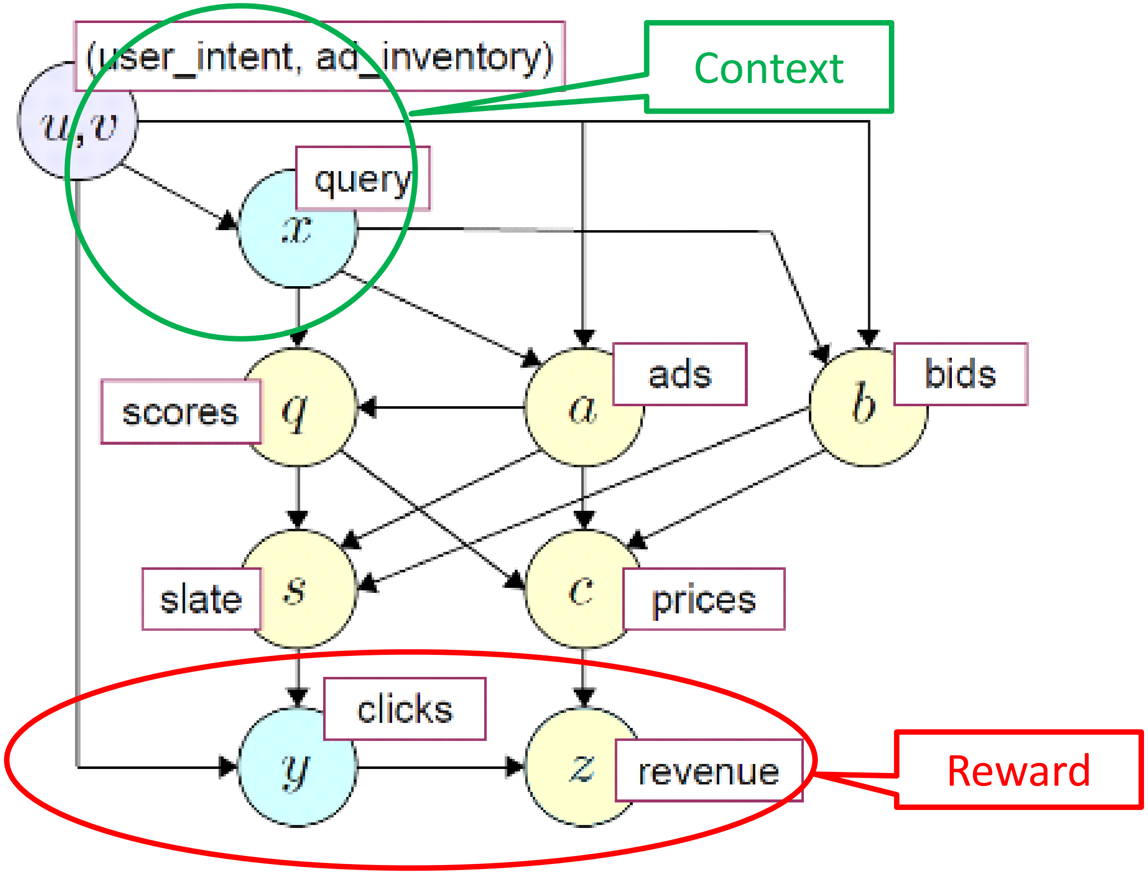
- They are interrelated because they share some factors.
- More complex algebraic interventions are of course possible.

A Contextual Bandit?



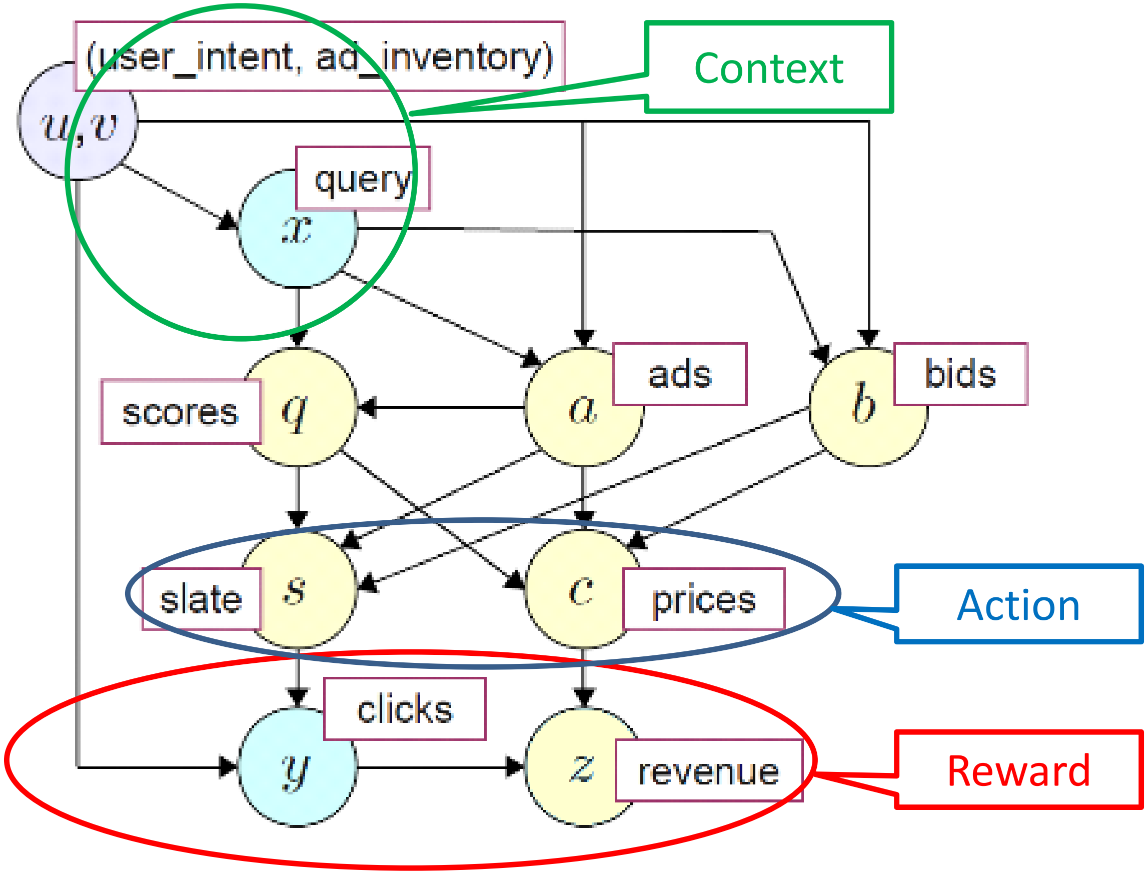
A Contextual Bandit?

- Context and potential rewards are drawn from joint unknown distribution
- Potential reward is a vector of rewards for all possible actions



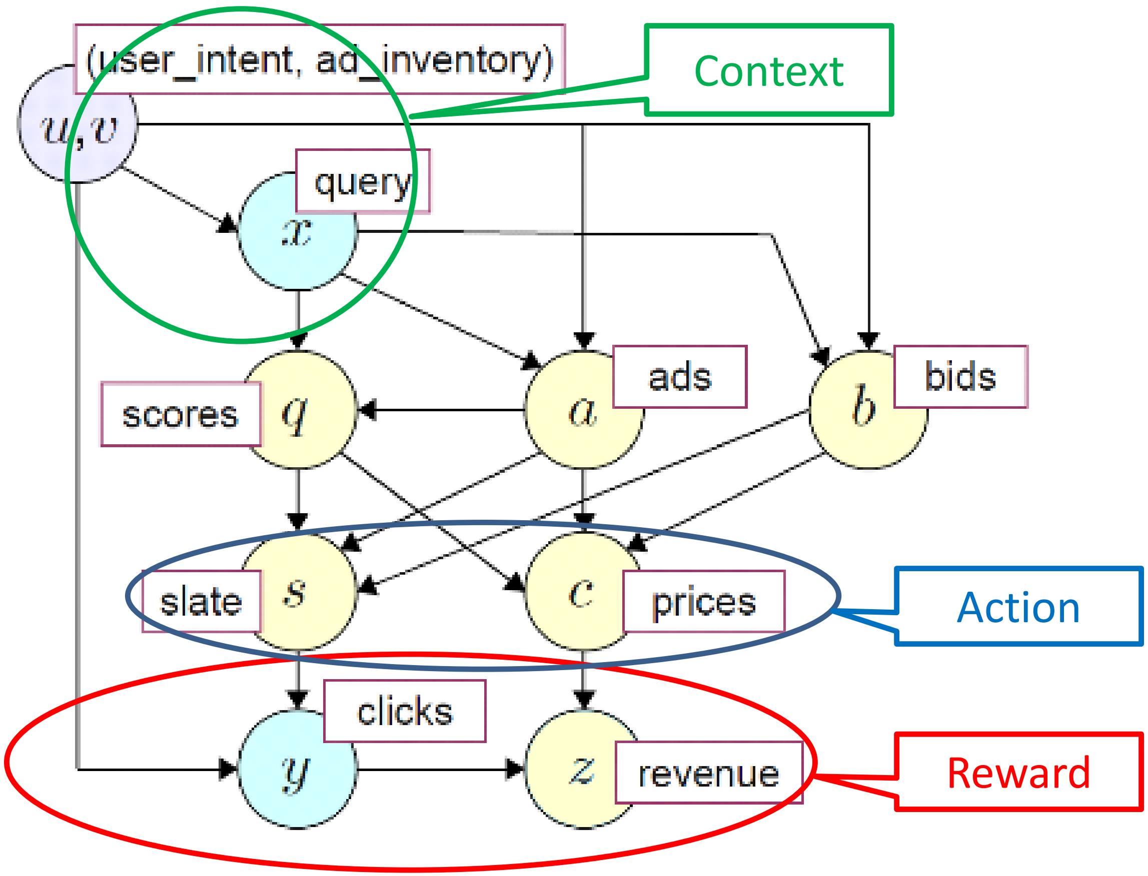
A Contextual Bandit?

- Context and potential rewards are drawn from joint unknown distribution
- Potential reward is a vector of rewards for all possible actions
- Action is decided by policy given context



A Contextual Bandit?

- Context and potential rewards are drawn from joint unknown distribution
- Potential reward is a vector of rewards for all possible actions
- Action is decided by policy given context



- A contextual bandit indeed
  - Very large context and action space
  - Structure in context, reward and policy

Importance sampling

$$
\begin{aligned}
\mathrm{P}^{*}(\omega) = \ & \mathrm{P}(u, v) \\
& \times \mathrm{P}(x \mid u) \\
& \times \mathrm{P}(a \mid x, v) \\
& \times \mathrm{P}(b \mid x, v) \\
& \times \cancel{\mathrm{P}(q \mid x, a)} \qquad \mathrm{P}^{*}(q \mid x, a) \\
& \times \mathrm{P}(s \mid a, q, b) \\
& \times \mathrm{P}(c \mid a, q, b) \\
& \times \mathrm{P}(y \mid s, u) \\
& \times \mathrm{P}(z \mid y, c)
\end{aligned}
$$

Distribution under intervention

- Can we estimate the results of the intervention counterfactually (without actually performing the intervention)
  - Yes if P and P* are non-deterministic (and close enough)

Importance sampling

Actual expectation

$$Y = \int_{\omega} \ell(\omega)\, P(\omega)$$

Counterfactual expectation

$$Y^* = \int_{\omega} \ell(\omega)\, P^*(\omega) = \int_{\omega} \ell(\omega)\, \frac{P^*(\omega)}{P(\omega)} P(\omega)$$

$$\approx \frac{1}{n} \sum_{i=1}^{n} \frac{P^*(\omega_i)}{P(\omega_i)}\, \ell(\omega_i)$$

Importance sampling

Principle

- Reweight past examples to emulate the probability they would have had under the counterfactual distribution.

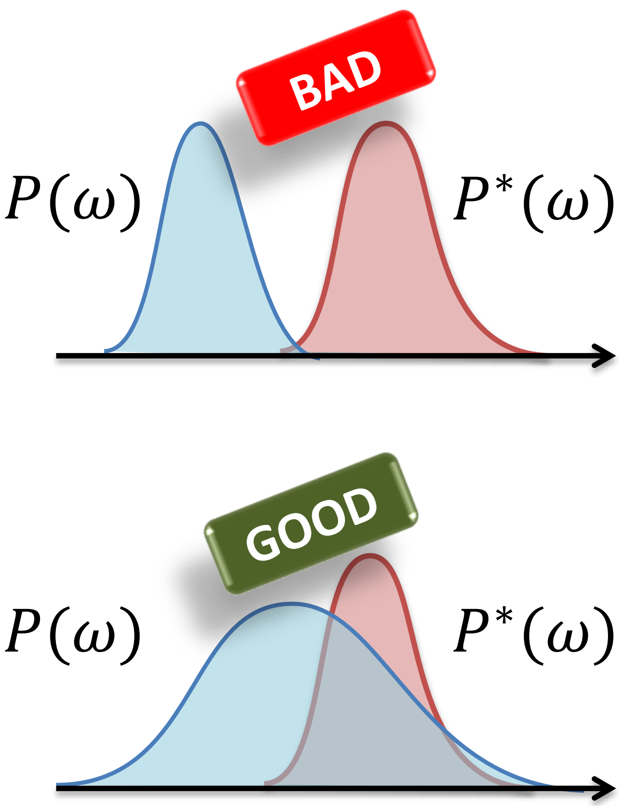
$$w(\omega_i) = \frac{P^*(\omega_i)}{P(\omega_i)} = \frac{P^*(q|x,a)}{P(q|x,a)}$$

Factors in P* not in P

Factors in P not in P*

- Only requires the knowledge of the factor under intervention (before and after)

Exploration



Quality of the estimation

- Large ratios undermine estimation quality.
- Confidence intervals reveal whether the data collection distribution $P(\omega)$ performs sufficient exploration to answer the counterfactual question of interest.

Confidence intervals

$$Y^* = \int_\omega \ell(\omega)\, w(\omega) P(\omega) \approx \frac{1}{n}\sum_{i=1}^{n} \ell(\omega_i)\, w(\omega_i)$$

Using the central limit theorem?

- $w(\omega_i)$ very large when $P(\omega_i)$ small.
- A few samples in poorly explored regions dominate the sum with their noisy contributions.
- Solution: ignore them.

Confidence intervals (ii)

Zero-clipped weights

$$\bar{w}(\omega) = \begin{cases} w(\omega) & \text{if less than } R, \\ 0 & \text{otherwise.} \end{cases}$$

Easier estimate

$$\bar{Y}^* = \int_\omega \ell(\omega)\, \bar{w}(\omega) P(\omega) \approx \frac{1}{n} \sum_{i=1}^{n} \ell(\omega_i) \bar{w}(\omega_i)$$

Confidence intervals (iii)

Bounding the bias

- Observe $\int_\omega w(\omega)P(\omega) = \int_\omega \frac{P^*(\omega)}{P(\omega)} P(\omega) = 1.$
- Assuming $0 \leq \ell(\omega) \leq M$ we have

$$0 \leq Y^* - \bar{Y}^* = \int_\omega [w - \bar{w}]\ell(\omega)\, P(\omega) \leq M \int_\omega [w - \bar{w}] P(\omega)$$

$$= M\left[1 - \int_\omega \bar{w}(\omega)P(\omega)\right] \approx M\left[1 - \frac{1}{n}\sum_{i=1}^{n} \bar{w}(\omega_i)\right]$$

- This is easy to estimate because $\bar{w}(\omega)$ is bounded.
- This represents what we miss because of insufficient exploration.

Two-parts confidence interval

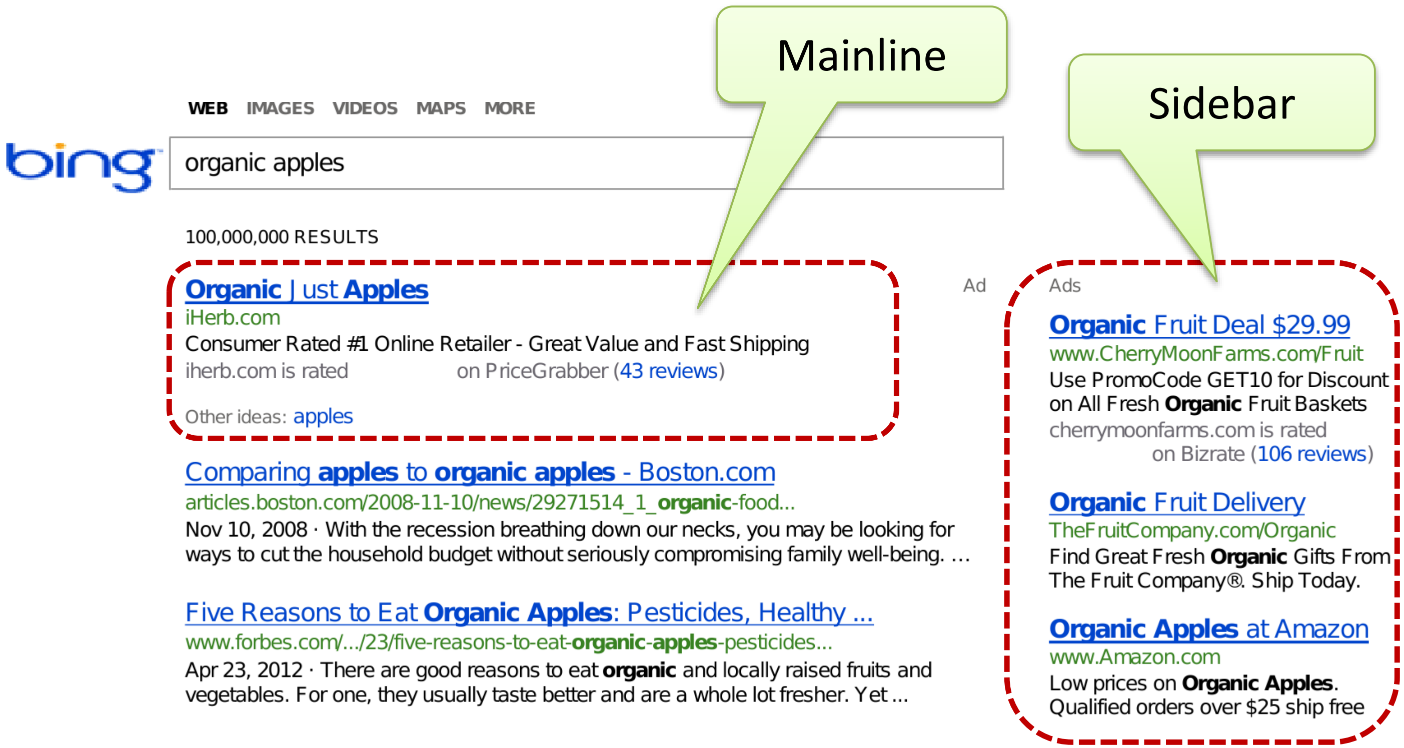
Outer confidence interval

- Bounds $\overline{\mathrm{Y}}^* - \overline{\mathrm{Y}}_n^*$
- When this is too large, we must sample more.

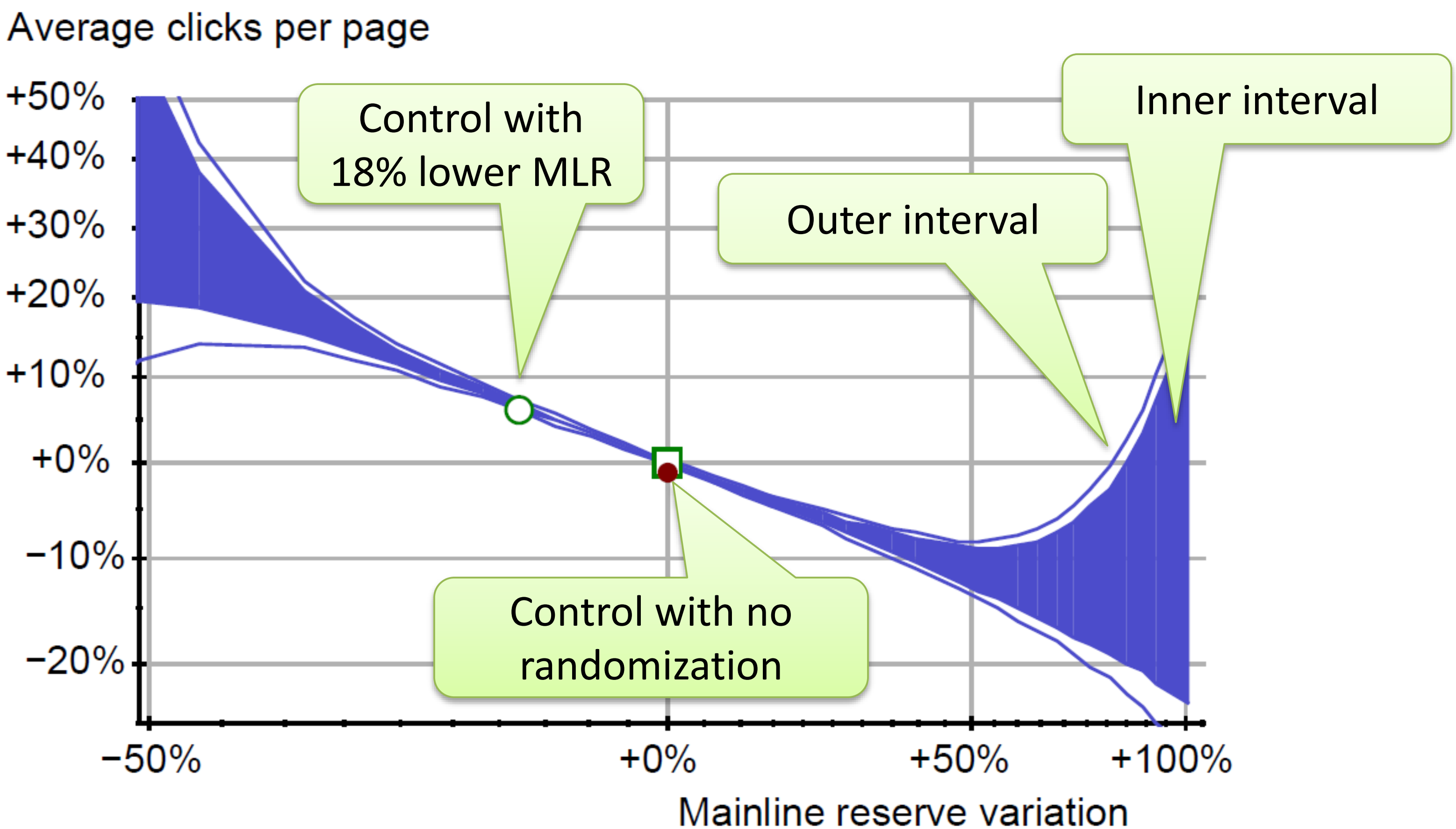
Inner confidence interval

- Bounds $Y^* - \bar{Y}^*$
- When this is too large, we must explore more.

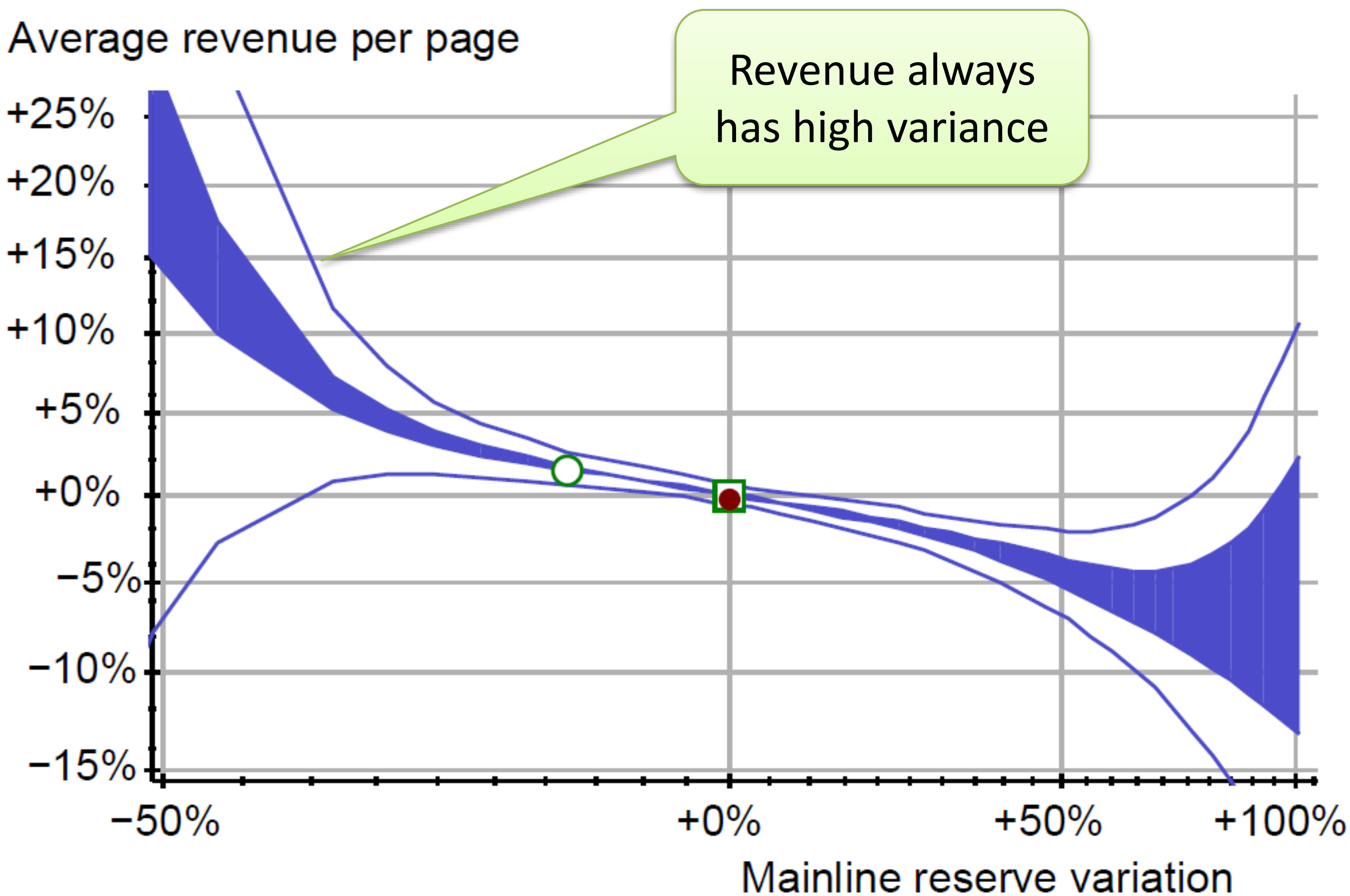
The pesky little ads again



Playing with mainline reserves (ii)



Playing with mainline reserves (iv)

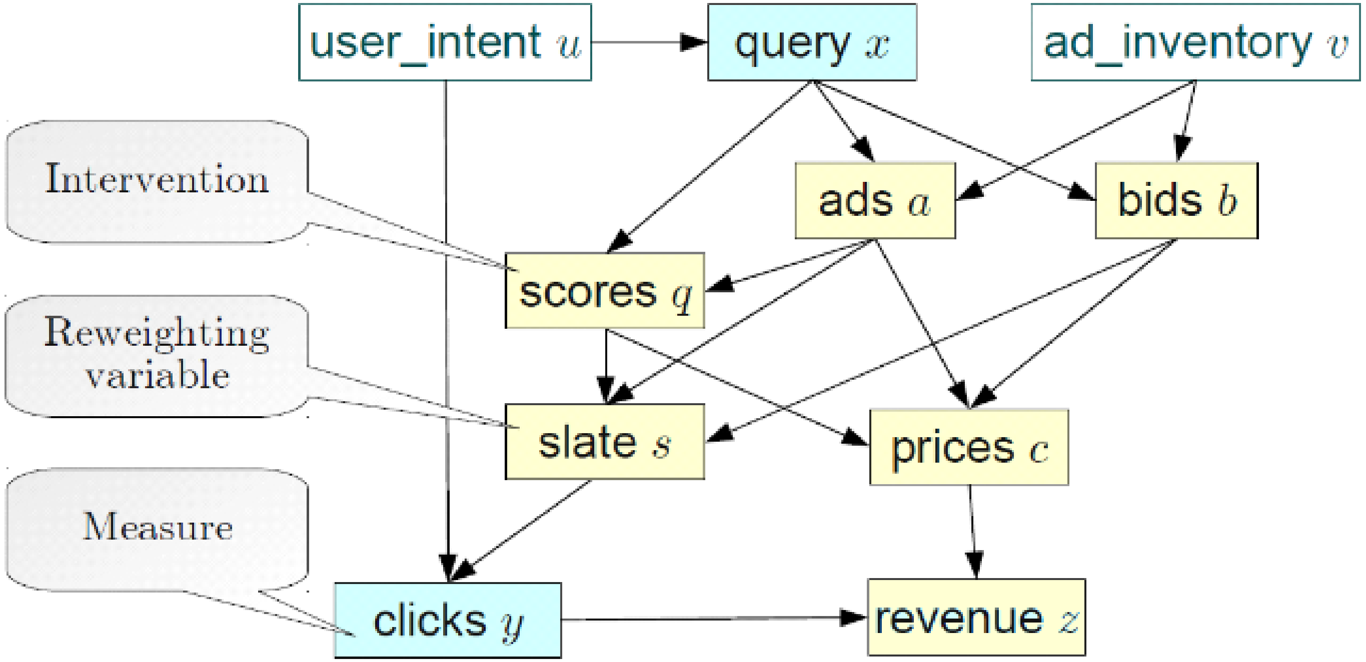


Algorithmic toolbox

- Improving the confidence intervals:
  - Exploiting causal graph for much better behaved weights
  - Incorporating neutral predictors invariant to the manipulation
- Counterfactual derivatives and optimization
  - Counterfactual differences
  - Counterfactual derivatives
  - Policy gradients
  - Optimization (= learning)
- Equilibrium analysis

Shifting the reweighting point

- Users make click decisions on the basis of what they see.
- They cannot see the scores, the reserves, the prices, etc.



Shifting the reweighting point

Standard weights

$$w(\omega_i) = \frac{P^*(\omega_i)}{P(\omega_i)} = \frac{P^*(q|x,a)}{P(q|x,a)}$$

Shifted weights

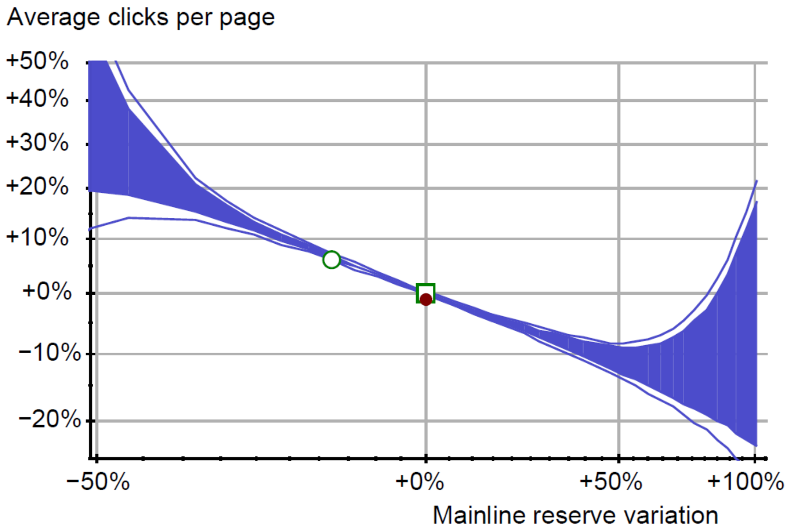
$$w(\omega_i) = \frac{P^*(s|x,a,b)}{P(s|x,a,b)}$$

with $P^\diamond(s|x,a,b) = \int_q P(s|a,q,b) P^\diamond(q|x,a)$.

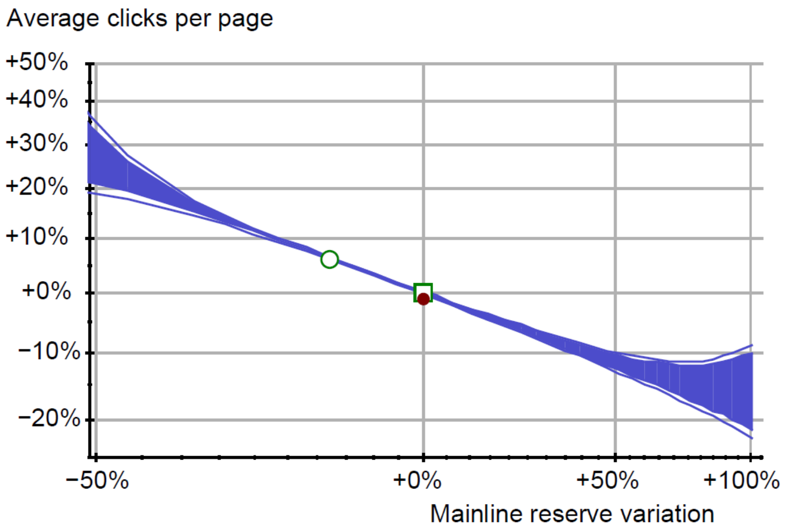
Shifting the reweighting point

Experimental validation

- Mainline reserves



Score reweighting



Slate reweighting

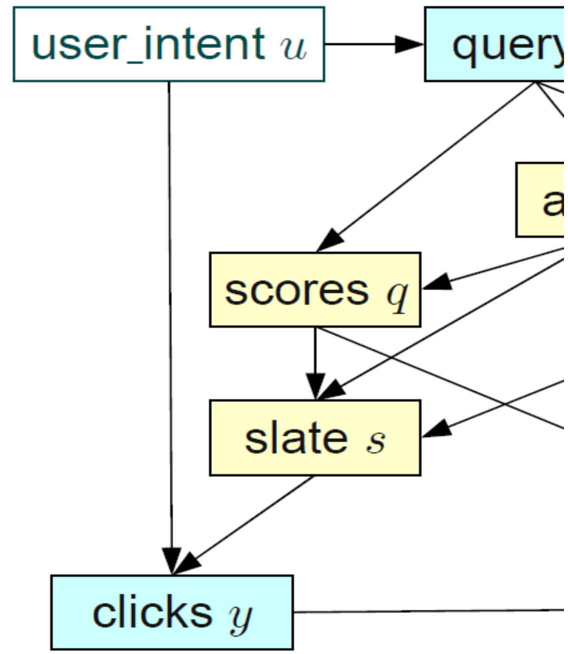
Shifting the reweighting point

When can we do this?

- $P^{\diamond}(\omega)$ factorizes in the right way iff
  1. Reweighting variables intercept every causal path connecting the point(s) of intervention to the point of measurement.
  2. All functional dependencies between the point(s) of intervention and the reweighting variables are known.

Shifting the reweighting point

- The engineering challenge:
  - The factor calculating slate based on scores is complex code
  - Need some way to calculate $P(s|x, a, b)$ automatically
- The organizational challenge:
  - Everybody wants to change the slate post-hoc



Shifting the reweighting point

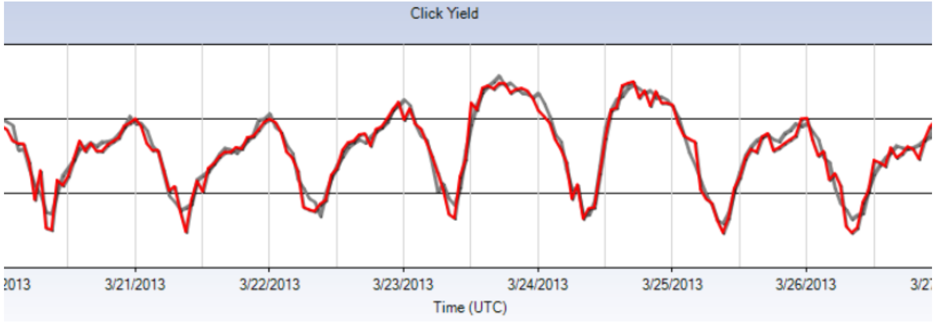
- The engineering challenge:
  - The factor calculating slate based on scores is complex code
  - Need some way to calculate $P(s|x, a, b)$ automatically

  - Code path leading to the slate as the reweighting variable
  - Symbolic algebra tracking the conditions leading to this code path
- The organizational challenge:
  - Everybody wants to change the slate post-hoc

  - Limit the information over which they can base their changes

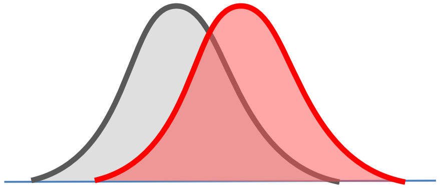
Variance reduction using a neutral predictor

Hourly average click yield for two interventions



$$\left(Y - \frac{1}{n}\sum y_i\right) \sim \mathcal{N}\left(0, \frac{\sigma}{\sqrt{n}}\right)$$

Daily effects increases the variance of both interventions.



Daily effects affect both interventions in similar ways!
Can we subtract them?

Variance reduction using a neutral predictor

- **Which intervention works best?**
  - Comparing expectations under counterfactual distributions $P^+(\omega)$ and $P^*(\omega)$.
  - Predictor $\zeta(x)$ estimates target on the basis of only the query time x.

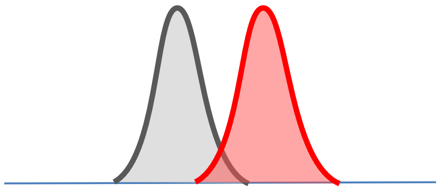
$$Y^+ - Y^* = \int_\omega [\ell(\omega) - \zeta(\nu)]\, \Delta w(\omega)\ P(\omega)$$

$$\approx \frac{1}{n}\sum_{i=1}^{n} [\ell(\omega_i) - \zeta(\nu_i)]\, \Delta w(\omega_i)$$

Variance captured by predictor $\zeta(\nu)$ is gone!

with $\Delta w(\omega) = \frac{P^+(\omega)}{P(\omega)} - \frac{P^*(\omega)}{P(\omega)}$

This is true regardless of the predictor quality.
But if it is any good, $\mathrm{var}[Y - \zeta(X)] < \mathrm{var}[Y]$, and



Main messages

- There are systems in the real world that are too complex to easily formalize
  - ML can assist humans in running these systems
- Relation between explore-exploit and correlation-causation
- The counterfactual framework provides a rich and modular framework for engineering of web-scale interactive learning systems