Computational Advertising & Causality

LÉON BOTTOU

MICROSOFT RESEARCH

UAI 2013

Joint work with

Jonas Peters

Denis Xavier Charles

Elon Portugaly

Patrice Simard

Joaquin Quiñonero Candela

D. Max Chickering

Dipankar Ray

Ed Snelson

1. Contents

Computational advertising ?

Online ad placement is a representative example of web-scale interactive machine learning system.

- Search engines.
- *Ad placement engines.*
- Recommendation systems.
- E-commerce systems.
- …

Scientific issues

Lots of relevant scientific literature

- Learning with limited feedback
  → multi-armed bandits, contextual bandits.
- Learning when actions change the environment
  → reinforcement learning.
- Independent decision making and incentives
  → auction theory.

From science to practice

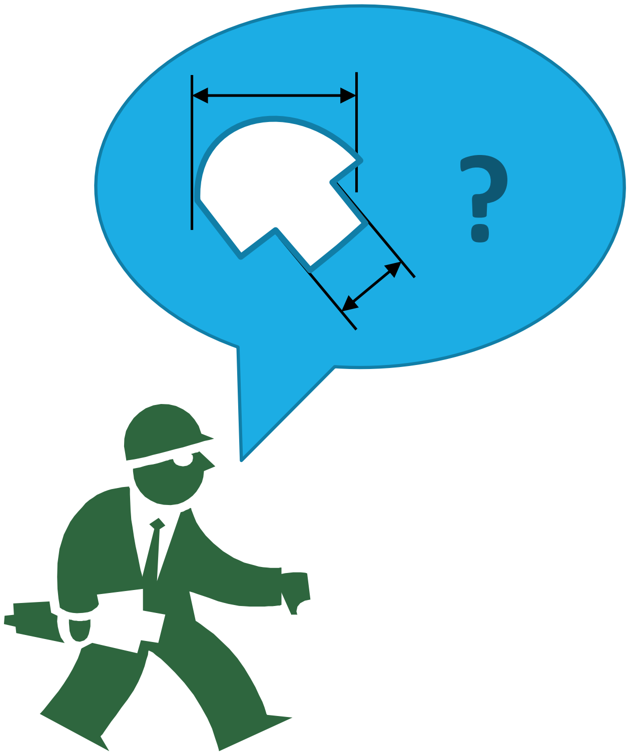
Sound scientific approach

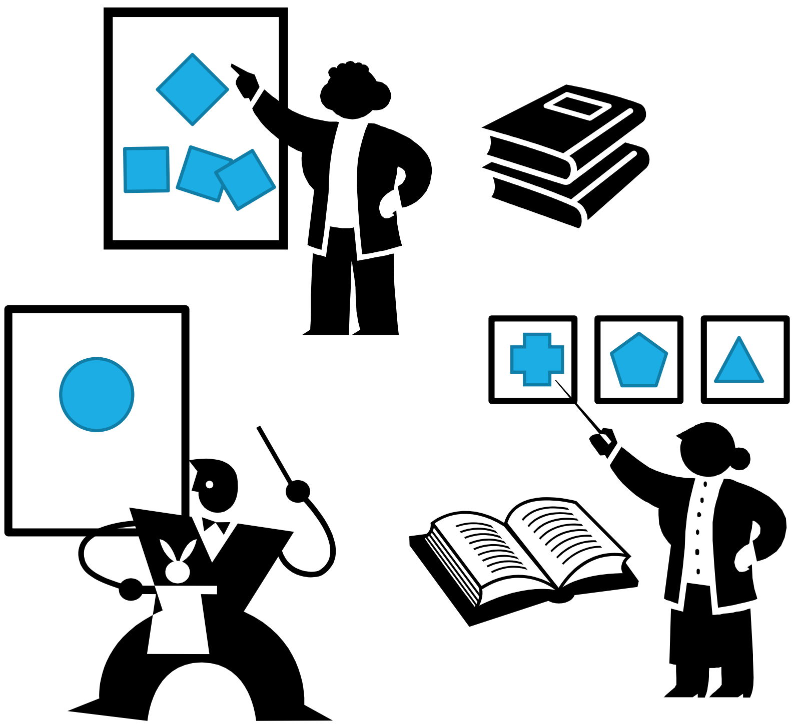
- Focus on the simplest setup that exhibits the phenomenon of interest and is amenable to analysis.

Practical consequences

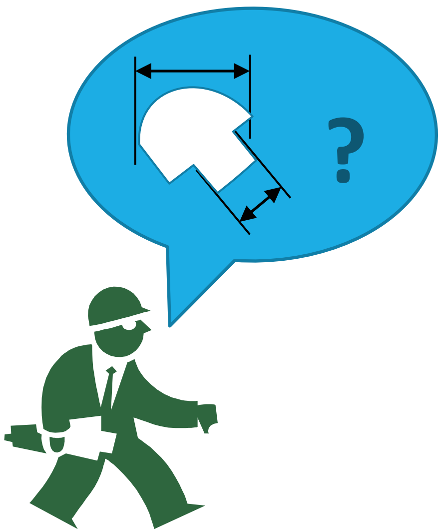
- Setup is too restrictive to apply to real systems.
- Setup is too general to lead to competitive solutions.
- Or both of the above.

Science is not about solutions





Reformulating the question



How to reason about my problem?

- Which language should I use to express the assumptions that I believe adequate for the problem.
- Which methods should I use to construct sound learning algorithms tailored to my assumptions.
- Which methods should I use to construct sound monitoring techniques to validate my assumptions, check the learning process at any time, debug new problems as they occur, etc.

This tutorial

Our answer

- Use the language and the methods of causal inference.

Tutorial

- Using ad placement as an example,
  describe a selection of useful causal inference methods,
  that form a good framework to solve the problem.

The rest of the story

Background

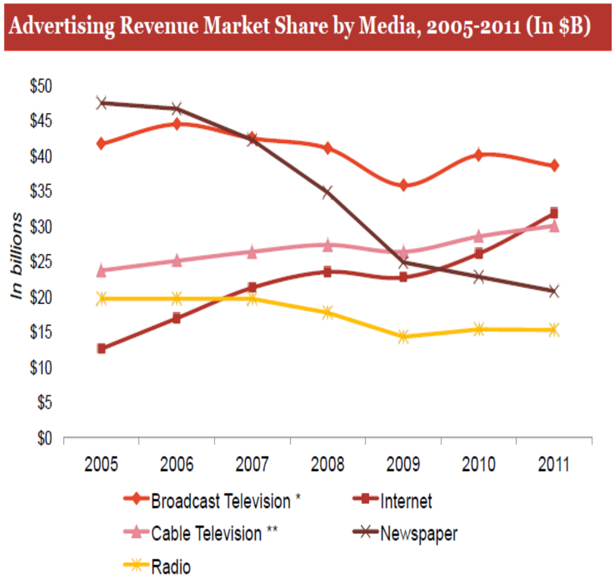
2. Ad Placement
3. Causation

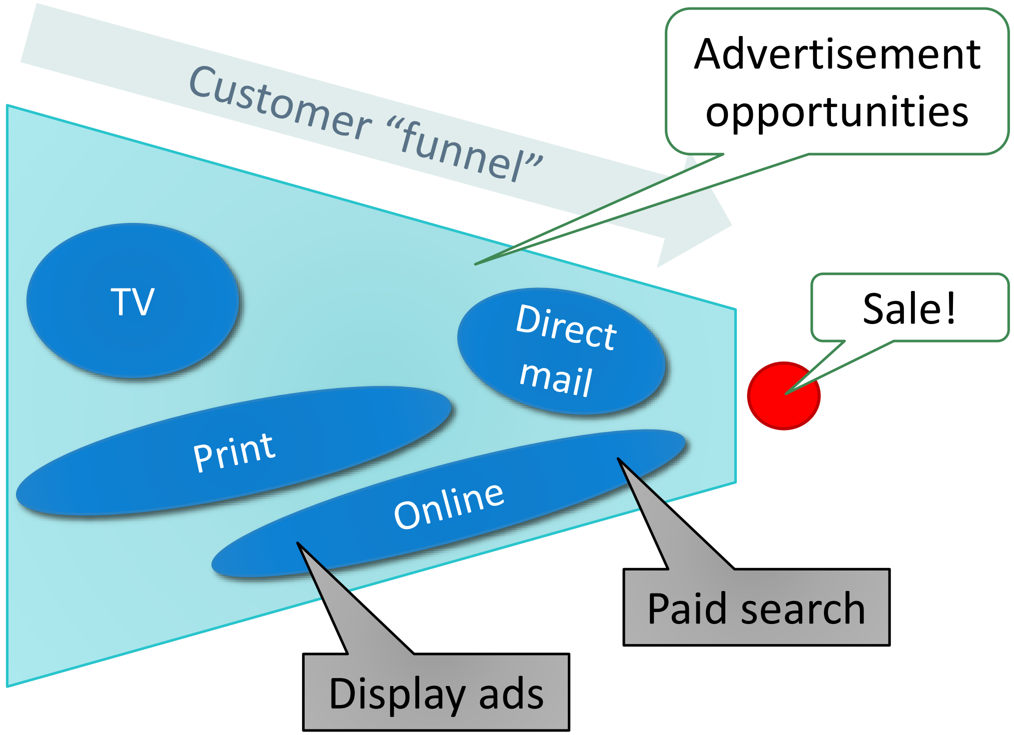
Methods

4. Counterfactual Evaluation
5. Leveraging the Causal Structure
6. Learning
7. Equilibrium

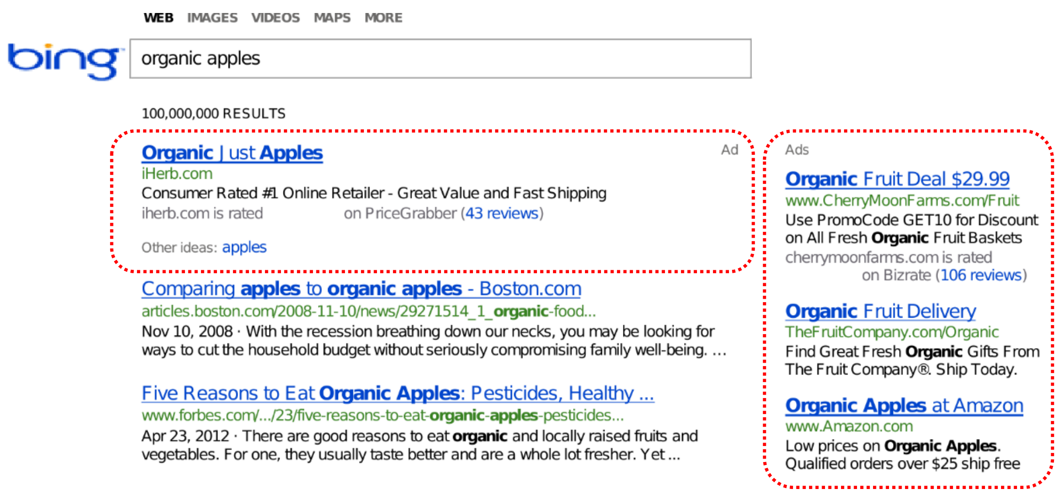
2. Ad Placement

Advertisement primer



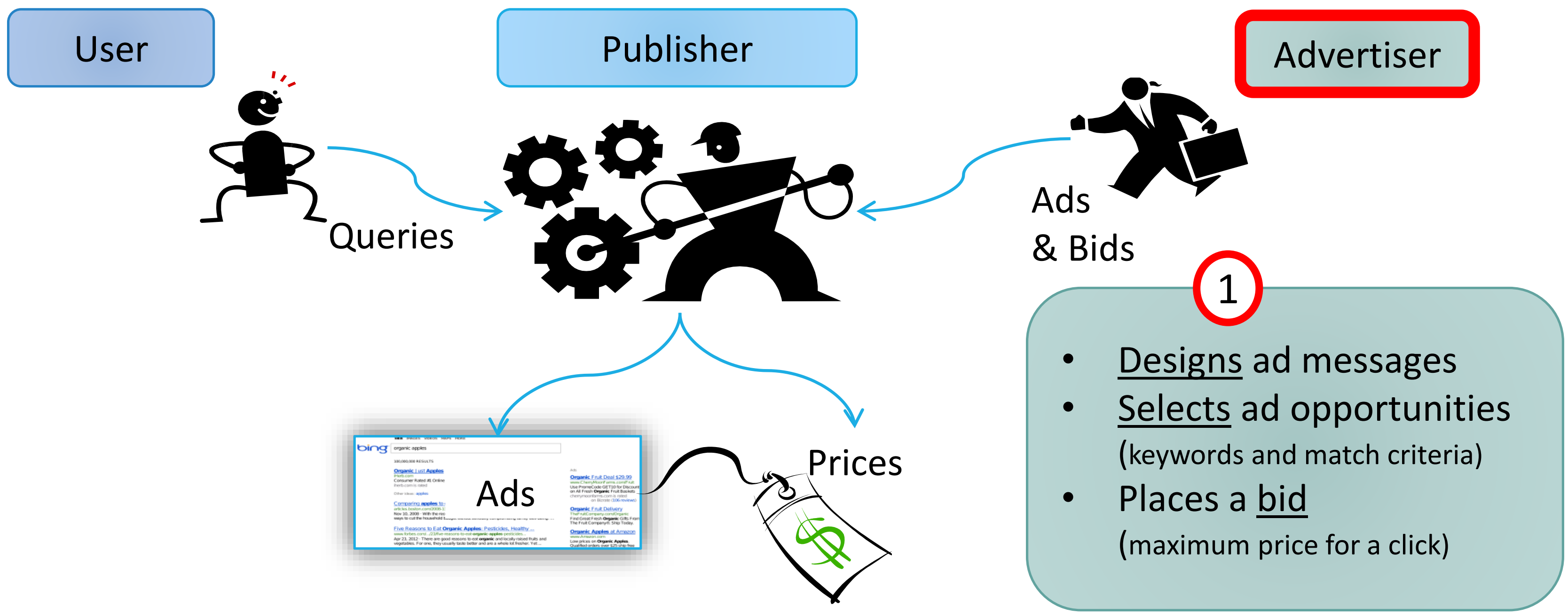


Paid search



- The most “effective” online ads are those displayed on search engines.
- How to choose which ads to display where?

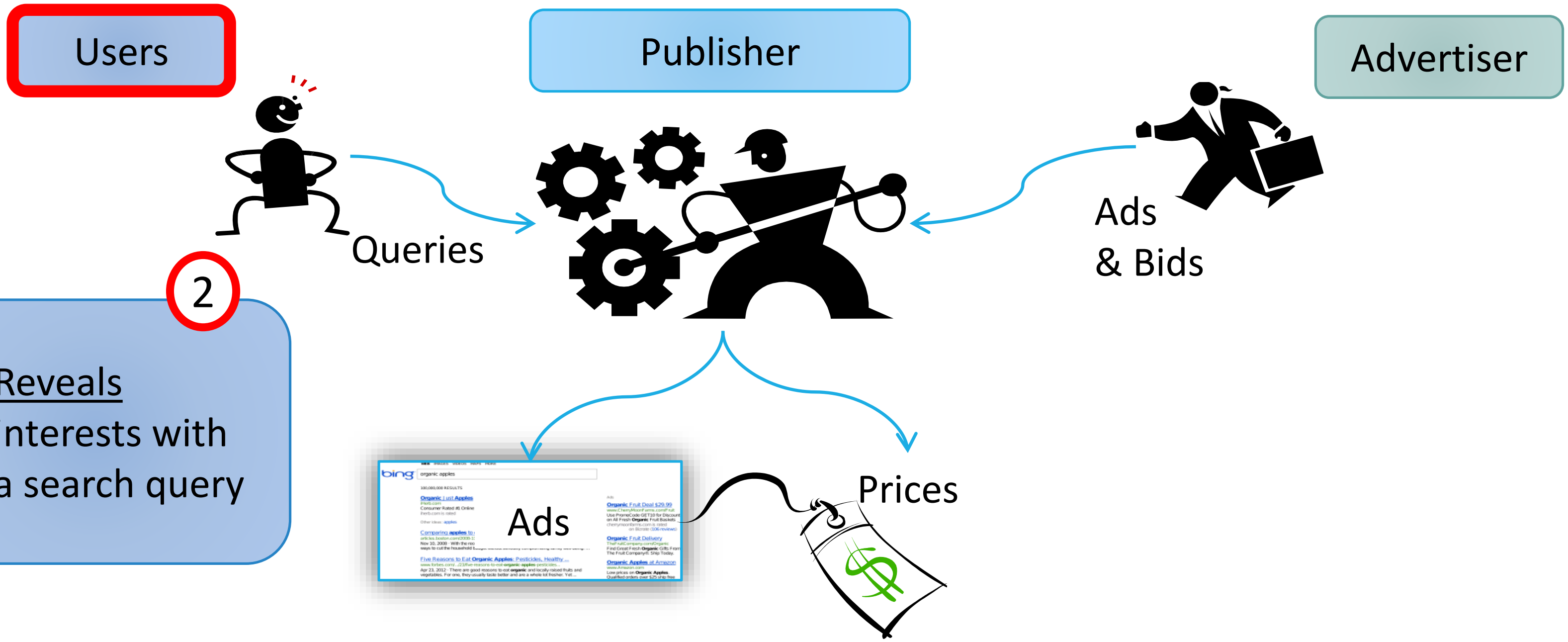
The game



In general, Paid search advertisers pay when the user clicks on their ad.

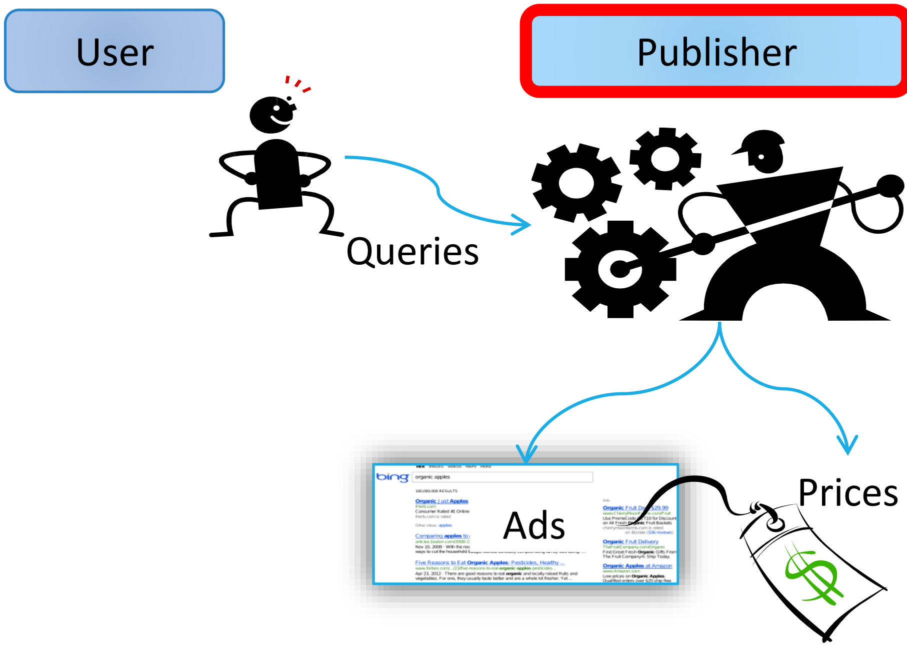
(there are other payment models, per impression, per action, etc.)

The game



- Reveals interests with a search query

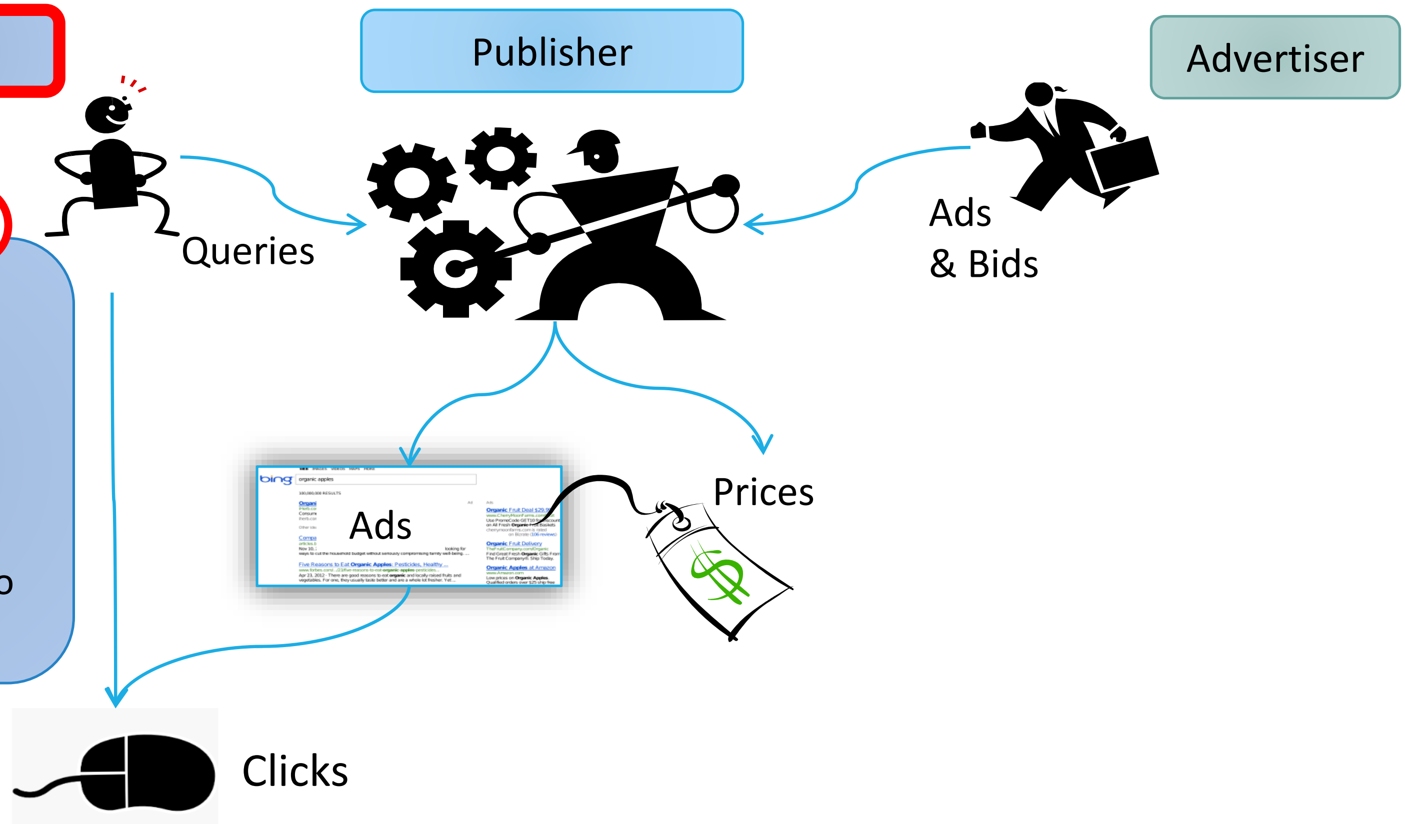
The game



3

- Computes search results
- Determines which ads to display (and where!)
- Determines price per click

The game

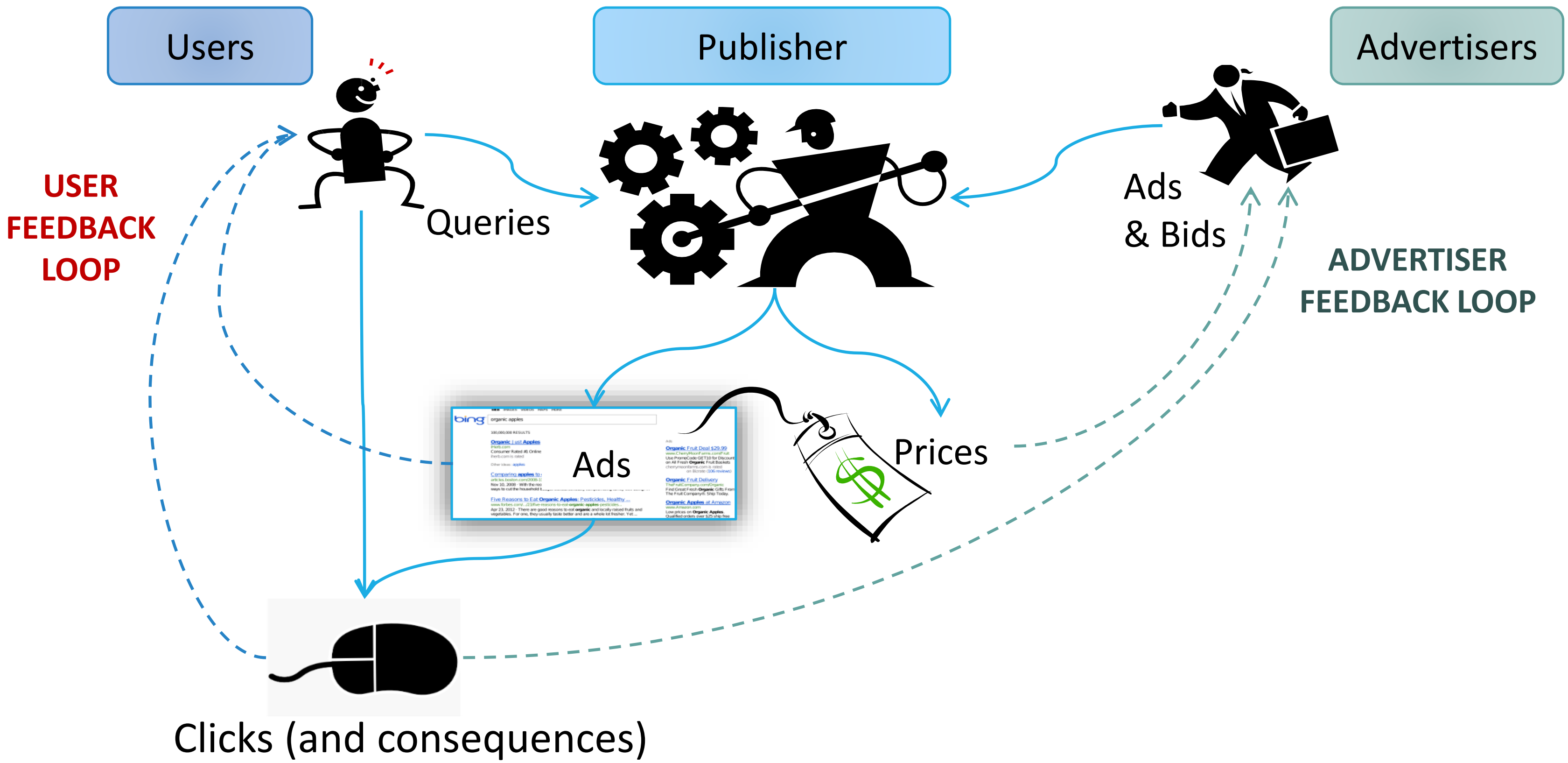


- May click on a relevant ad and jump to the advertiser site
- Triggering a payment from the advertiser to the publisher.

Self interest

User	• <u>Expects</u> results that satisfy her interests • <u>Possibly</u> by initiating business with an advertiser • Future engagement depends on her satisfaction….
Advertiser	• <u>Expects</u> to receive potential customers • Expects to <u>recover</u> clicks costs from resulting business • Return on investment impacts future ads and bids…
Publisher	• <u>Expects</u> click money • <u>Learns</u> which ads work from past data. • In order to <u>preserve</u> future gains, publishers must ensure the continued satisfaction of users and advertisers. (this changes everything!)

Second order effects



Auctions

Setup

- Seller has an object to sell.
- Each buyer values the object differently.
- Each buyer knows the auction mechanism and places a bid.
- Auction mechanism determines who gets the object and how much he pays.

Notes

- The auction outcomes are functions of the bids.
- Buyer bids according to his value and his beliefs about other buyers values.
- The value of whoever gets the object is the size of the pie.
- The payment from the buyer to the seller then splits the pie.

Which mechanism works best for the seller?

Auctions

A first auction mechanism

"The highest bidder receives the object and pays his bid."

- Buyers should bid less than their value.
  - If they bid their value, their surplus is zero in all cases.
  - If they bid more, they may get the object with a negative surplus.
  - If they bid less, they trade a chance to lose the object for a chance to pay less.

- The object may not go to the buyer who values it most.

→The expected pie is smaller and the expected buyer surplus is larger.
This cannot be good for the seller.

- The object may sell for less than the seller's value.

→ Can use a reserve price, that is, an additional bid entered by the seller.

Auctions

A second auction mechanism

"The highest bidder receives the object and pays the second highest bid."

- Buyers now should bid their value ("*truthful mechanism*")
  - Overbidding buyers may get the object with negative surplus.
  - Underbidding buyers will not pay less if they get the object. On the other hand, they may see the object sold to another buyer for less than their value, losing the opportunity to have a positive surplus.

    Unless a buyer is certain that no other buyer will bid above a level smaller than his value, the buyer best interest is to bid his value, regardless of his exact beliefs.

- The object always goes to the buyer who values it most.
- The object may still sell for less than the seller's value.

Auctions

A third auction mechanism

"The seller announces a reserve price which works like an additional bid.
- If the highest bid is the reserve price, the seller keeps the object.
- Otherwise the highest bidder receives the object and pays the second highest bid."

- Buyers should still bid their value ("*truthful mechanism*")
- But the seller should set a reserve price that is higher than his value!
  - He trades the risk of not selling for the chance to get more than his value.
  - Therefore the object may not sell even though a buyer values it more than the seller. This in fact makes the pie smaller in a manner that benefits the seller.
- Under mild assumptions, this is the optimal mechanism for the seller.
  - For the correct value of the reserve price, of course…

Ad placement auctions

Mapping auction theory to ad placement

- The publisher is the seller (he receives bids)
- The advertisers are the buyers (they place bids)
- What about click decisions made by the user?
- What is the "object" exactly?

Click probabilities

The click probabilities $(q_1, \ldots, q_k)$ of the eligible ads $(a_1, \ldots, a_k)$ depend

- on the context x, that is, the query, the user, the session, the weather…
- on the ad messages themselves $(a_1, \ldots, a_k)$,
- on the positions $(p_1, \ldots, p_k)$ chosen by the publisher,
- but do not depend on the bids $(b_1, \ldots, b_k)$.

Ad placement auctions

One of the many ways to view ad placement auctions…

The auction mechanism specifies

- The probability that each competing advertiser gets a click (the object).
- The expected price paid by each competing advertiser.

There is an optimal mechanism (Myerson, 1981).

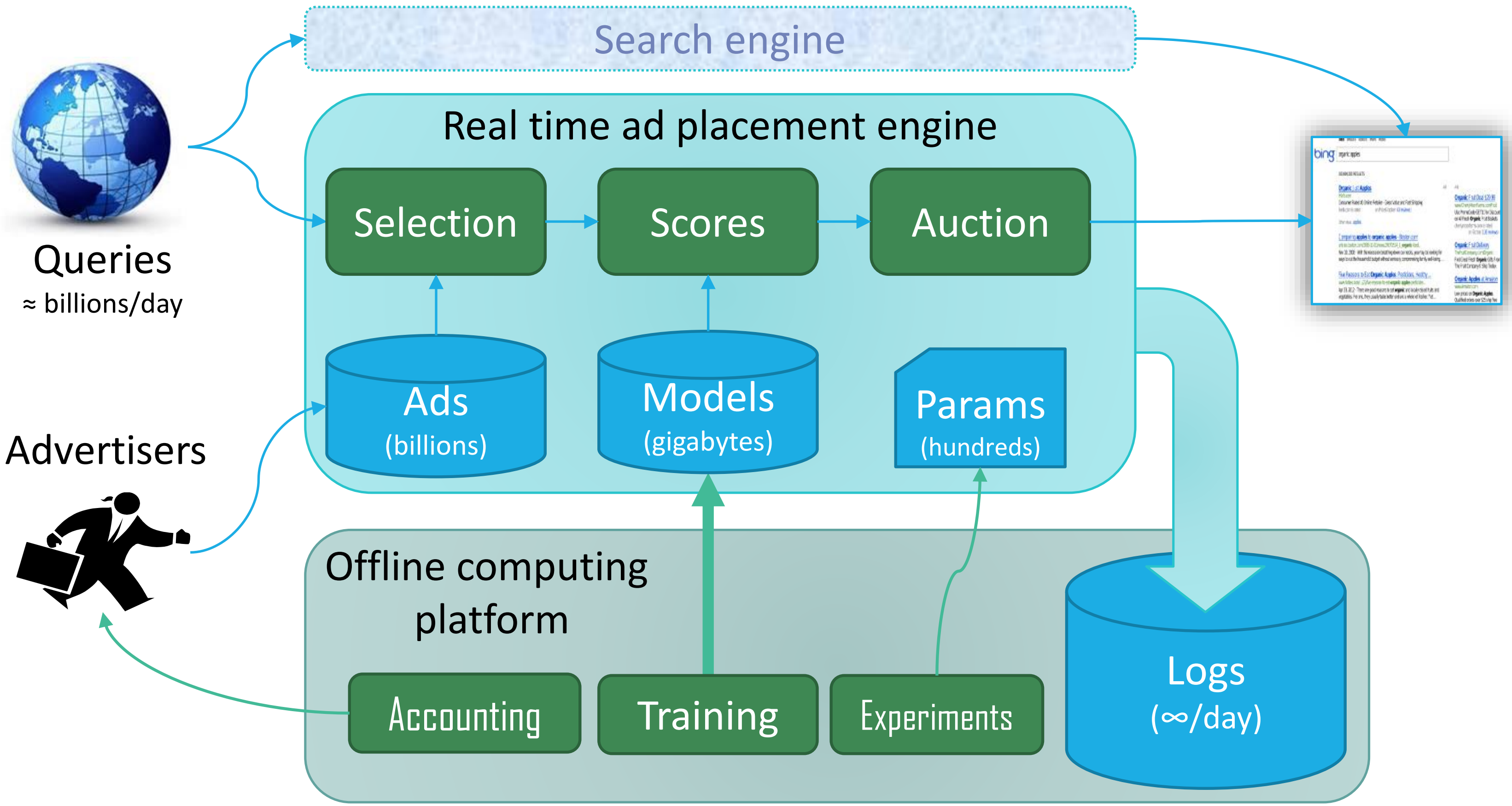
- The placement $(p_1, \ldots, p_k)$ maximizes $\sum_i b_i \times q_i(x, \boldsymbol{a}, \boldsymbol{p})$ subject to reserves.
- The prices are determined by the Vickrey-Clarke-Groves (VCG) rule,
  a nontrivial generalization of the second price rule.

- See also (Varian, 2007; Edelman et al., 2007)

Optimal auctions ?

- Many queries are targeted by a single advertiser.
  - When there is only one buyer, this is not an auction!
- The optimal auction theory is valid for a single auction.
  - The optimal auction might leave the buyer quite unhappy
    This is not going to work if we deal again and again with the same buyer…
- Advertisers place a single bid for multiple auctions.
  - An ad can be eligible for a lot of different queries.
  - The Bing/Yahoo engine serves hundreds of millions of queries per day.
    The most active advertisers change their bids every 15 minutes.
- Placement decisions impact the future behavior of users.
  - Some advertisers try to cheat the users by directing them to spam sites.
    This is not good for the long term revenue of the publisher.

How it really works



How it really works

**The following mechanism is the result of history.
This is what the advertisers expect. Changing it is hard!**

1. Publisher selects eligible ads $(a_1, \ldots, a_k)$ for the query x.
2. Publisher computes click scores q_i and rank scores r_i

$$q_i(x, a_i, p_i) = \underbrace{\gamma(x, p_i)}_{\text{Position effect}} \times \underbrace{\beta(x, a_i)}_{\text{Ad effect}} \qquad r_i(x, a_i) = \underbrace{b_i}_{\text{Bid}} \times \underbrace{\beta(x, a_i)}_{\text{Ad effect}}$$

3. Publisher greedily assigns ads with the largest rank scores to the best available positions, until reaching a predefined reserve score
4. Generalized second price (GSP): advertiser pays the smallest bid that would have guaranteed the same placement.

The ugly truth

2. Publisher computes click scores q_i and rank scores r_i

$$q_i(x, a_i, p_i) = \gamma(x, p_i) \times \beta(x, a_i) \qquad r_i(x, a_i) = b_i \times \beta(x, a_i)$$

No longer a pure click probability.
Secret ingredients attempt to represent user satisfaction.

4. Generalized second price.

The auction is not truthful because GSP is not VCG.

Furthermore, additional ingredients give discounts for certain auctions.

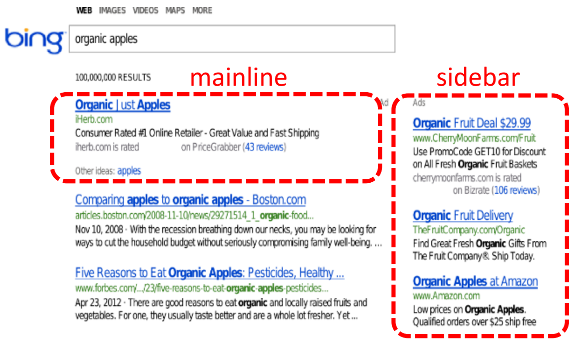
I do not understand the combined effects of all these adjustments.
I have never met anyone who could explain them to me.

Experimentation : A/B Testing

How to compare two ad placement engine variants?

1. Implement both variants
2. Randomly split traffic in two “buckets”
   - Place treatment buckets ads using the variant under investigation.
   - Place control buckets ads using the normal placement engine.
3. Run for some time and measure performance metrics.

Performance metrics



First order performance metrics

- Average number of ads shown per page
- Average number of **mainline ads** per page
- Average number of ad clicks per page
- Average revenue per [mille] page (RPM)

Should we just optimize RPM?

Showing lots of mainline ads improves RPM.
Users would quickly go away!

Increasing the reserve prices also improves RPM.
Advertisers would quickly go away!

Performance metrics

First order performance metrics

- Average number of ads shown per page
- Average number of mainline ads per page
- Average number of ad clicks per page
- Average revenue per page (RPM)
- Average relevance score estimated by human labelers
- Average number of bid-weighted ad clicks per page
- …

Monitor heuristic indicators of user fatigue

Monitor heuristic indicators of advertiser value

Splitting traffic

Long term user feedback experiments

Measure actual user fatigue instead of heuristic indicators.

- Randomly split users into treatment and control groups.
- Wait a couple months and compare performance metric.
- This comparison reveals second order user effects…

Long term advertiser feedback experiments

- Randomly split advertisers into treatment and control groups
- Which version of the ad placement engine should we run when an auction involves advertisers from both groups ?

Problems with A/B testing

No single decision criterion

- Because of complex second order effects.

Requires full implementation of treatment.

Must wait two weeks for significant results.

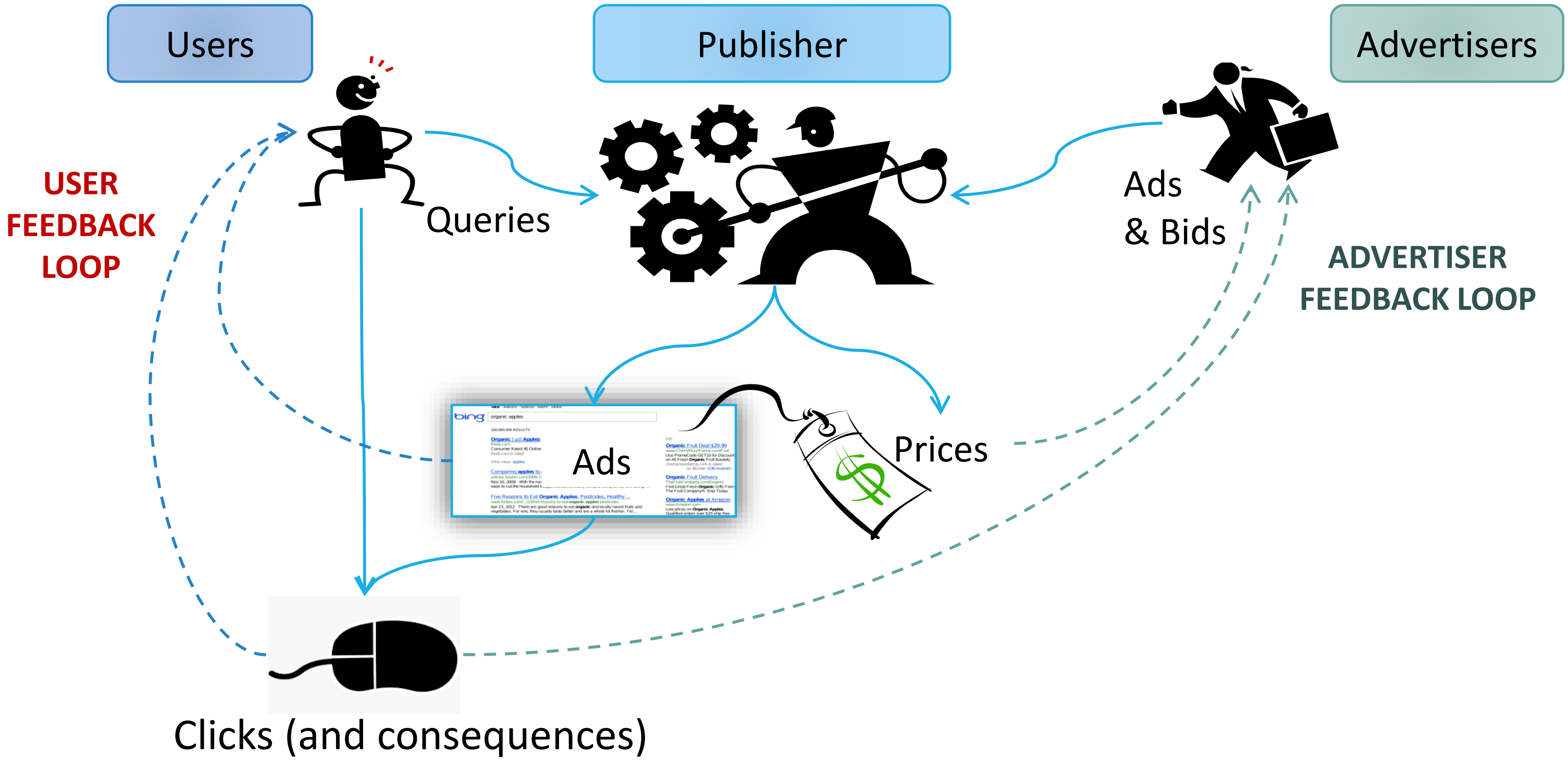
- Impractical for the early development of new ideas.
- Cannot drive learning algorithms.

Experimentation is limited by total traffic.

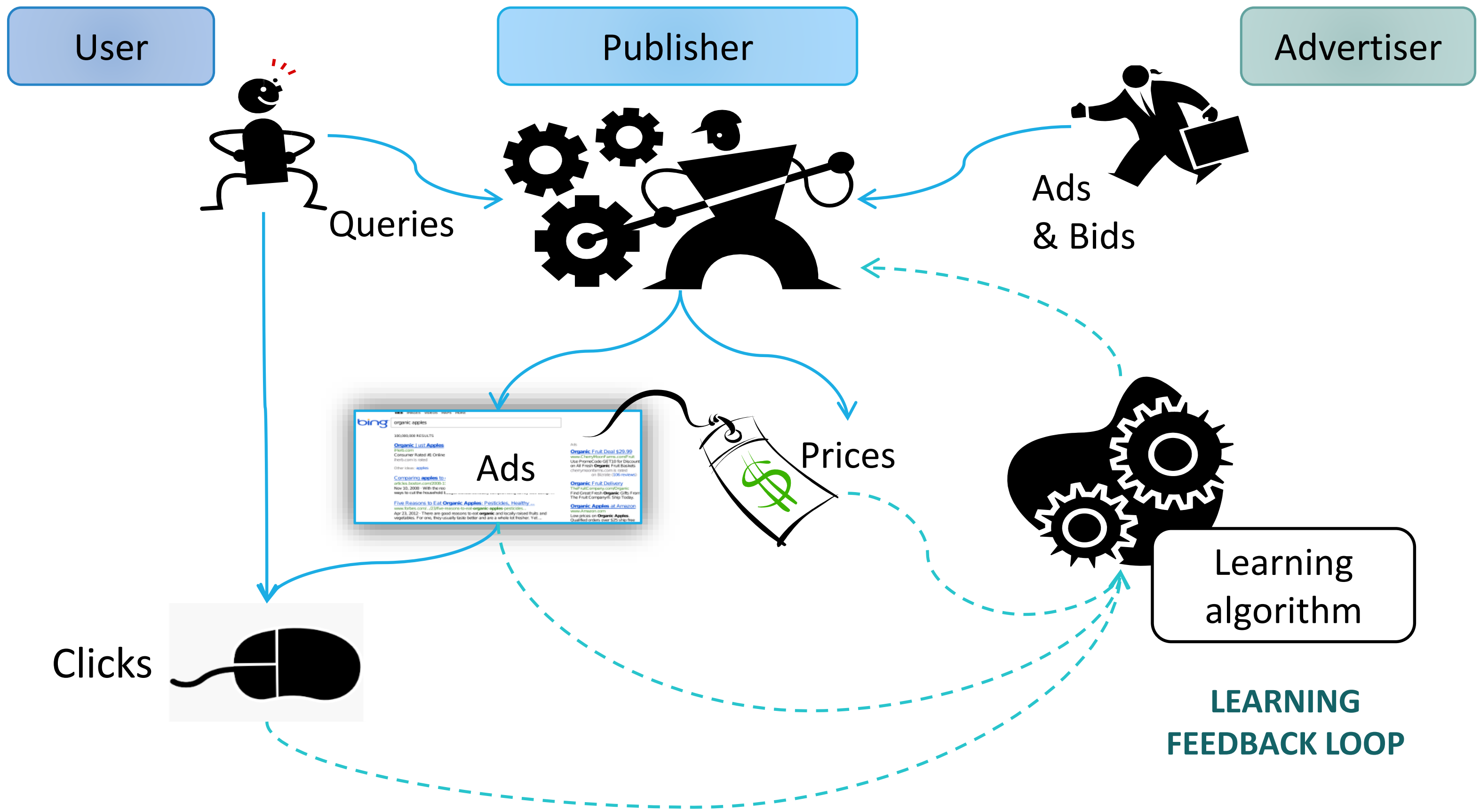
- Hundreds of experiments are running at the same time.
- Overlapped experiments.

2½. Feedback Loops

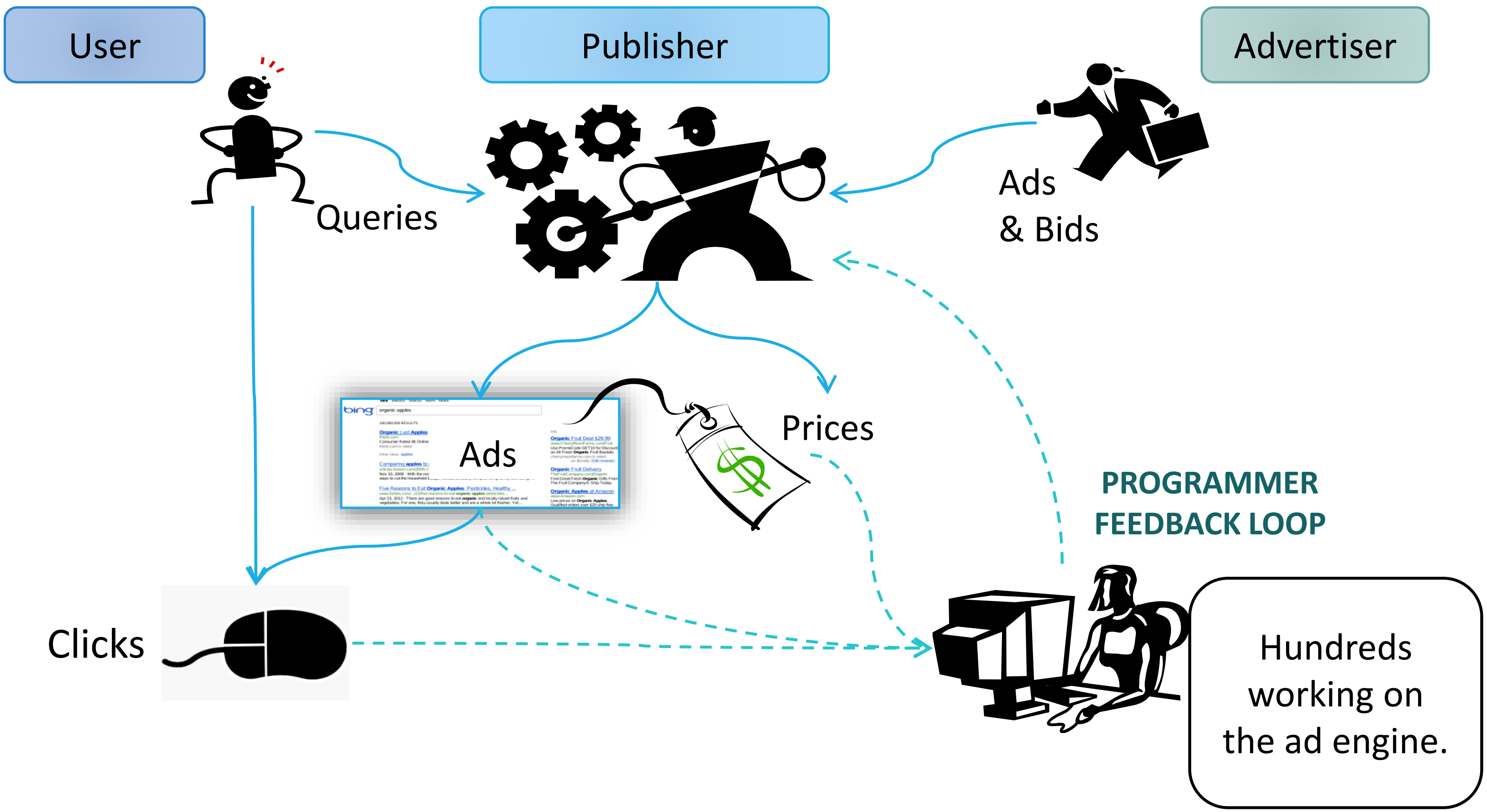
User and advertiser feedback



Learning feedback



Engineering feedback



The feedback loop problem

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

Two queries

Q1: “cheap diamonds” (50% traffic)
Q2: “google” (50% traffic)

Three ads

A1: “cheap jewelry”
A2: “cheap automobiles”
A3: “engagement rings”

More simplifications

- We show only one ad per query
- All bids are equal to $1.

Toy example

True conditional click probabilities

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

Step 1: pick ads randomly.

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

Toy example

Step 2: estimate click probabilities

- Build a model based on a single Boolean feature:
  F : “*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 (google)	2%	2%	2%

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

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

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 (google)	2%	2%	2%

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

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 (google)	2%	2%	2%

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

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

- 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 F identifies relevant ads using a narrow criterion.

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

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

3 P(Click|¬F) is pulled down by queries that do not click.

4 Ads for query Q1 are ranked incorrectly.

Adding a feature 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.

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

Adding a (Q1,A3) feature

- 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.

3. Causation

Causal paradoxes in ad data

A legitimate question

"Does it help to know the estimated click probability of the first mainline ad in order to estimate the click probability of the second mainline ad?"

Naïve approach

- Collect past data for pages showing at least two ads.
- Split them in two groups according to the estimated click probability q_1 computed for the first ad.
- Count clicks on the second ad and compare.

Causal paradoxes in ad data

	CTR_2
q_1 low	124/2000 (6.2%)
q_1 high	149/2000 (7.5%)

Confounding factors…

- Commercial queries get higher q_1.
  They also receive more clicks everywhere…
- Let us split the data according to the estimated click probability q_2 computed for the second ad.

Causal paradoxes in ad data

	CTR_2	if q_2 low	if q_2 high
q_1 low	124/2000 (6.2%)	92/1823 (5.1%)	32/176 (18.1%)
q_1 high	149/2000 (7.5%)	71/1500 (4.8%)	78/500 (15.6%)

Simpson reversal!

- This happens because q_1 and CTR_2 have a (confounding) common cause.
- What about the common causes we do not know?
- We cannot trust statistical correlations in collected data.
- Most of machine learning is about modeling correlations.

Statistics and Causation

Correlations have predictive value

- *“It is raining”* ⇒ *“People probably carry open umbrellas*.”
- *“People carry open umbrellas”* ⇒ *“It is probably raining.”*

Interventions

- Hypothetical: *“Will it rain if we ban umbrellas?”*
- Counterfactual: *“Would have it rained if we had banned umbrellas?”*

Causation

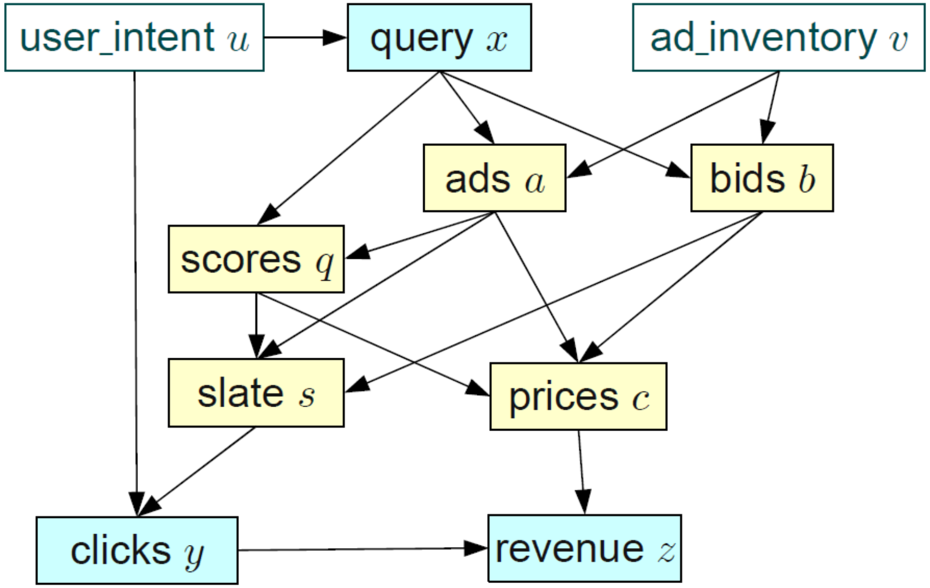
- Causal relations let us to reason on the outcome of interventions.

Recent advances in causal inference

- (Rubin, 1986) (Spirtes et al., 1993, 2011, …)
- (Pearl, 2000, 2009, …)

Structural equation model (SEM)

$$
\begin{aligned}
x &= f_1(u, \varepsilon_1) \\
a &= f_2(x, v, \varepsilon_2) \\
b &= f_3(x, v, \varepsilon_3) \\
q &= f_4(x, a, \varepsilon_4) \\
s &= f_5(a, q, b, \varepsilon_5) \\
c &= f_6(a, q, b, \varepsilon_6) \\
y &= f_7(s, u, \varepsilon_7) \\
z &= f_8(y, c, \varepsilon_8)
\end{aligned}
$$



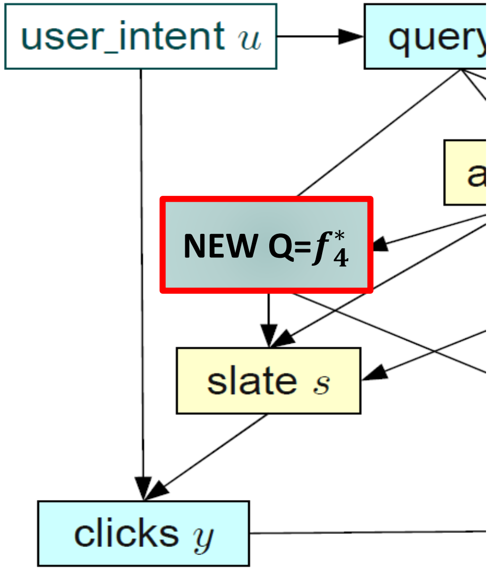
Direct causes / Known and unknown functions

Noise variables / Exogenous variables

(Wright, 1921)

Interventions

$$
\begin{aligned}
x &= f_1(u, \varepsilon_1) \\
a &= f_2(x, v, \varepsilon_2) \\
b &= f_3(x, v, \varepsilon_3) \\
q &= \cancel{f_4(x, a, \varepsilon_4)} \quad f_4^*(x, a, \varepsilon_4) \\
s &= f_5(a, q, b, \varepsilon_5) \\
c &= f_6(a, q, b, \varepsilon_6) \\
y &= f_7(s, u, \varepsilon_7) \\
z &= f_8(y, c, \varepsilon_8)
\end{aligned}
$$



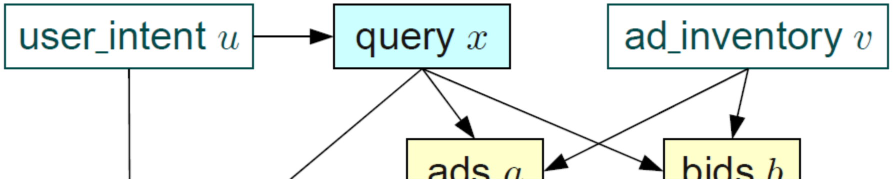
Interventions as algebraic manipulation of the SEM.
Causal graph must remain acyclic.

Isolation

What to do with unknown functions?

- Replace knowledge by statistics.
- Statistics need repeated isolated experiments.
- Isolate experiments by assuming an unknown but invariant joint distribution for the exogenous variables.

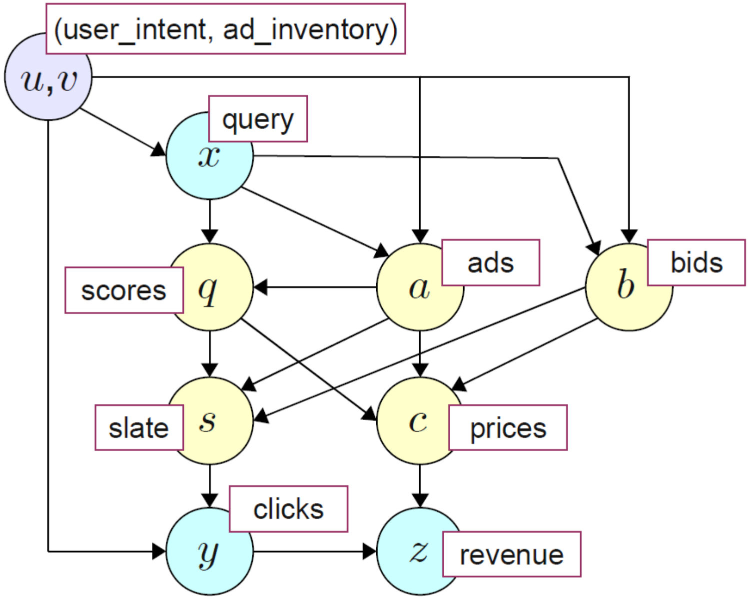
$$P(u, v)$$



⇒ No feedback loops (…yet…)

Markov factorization

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



This is a “Bayes network”

a.k.a. “directed acyclic probabilistic graphical model.”

(Pearl, 1988)

Markov interventions

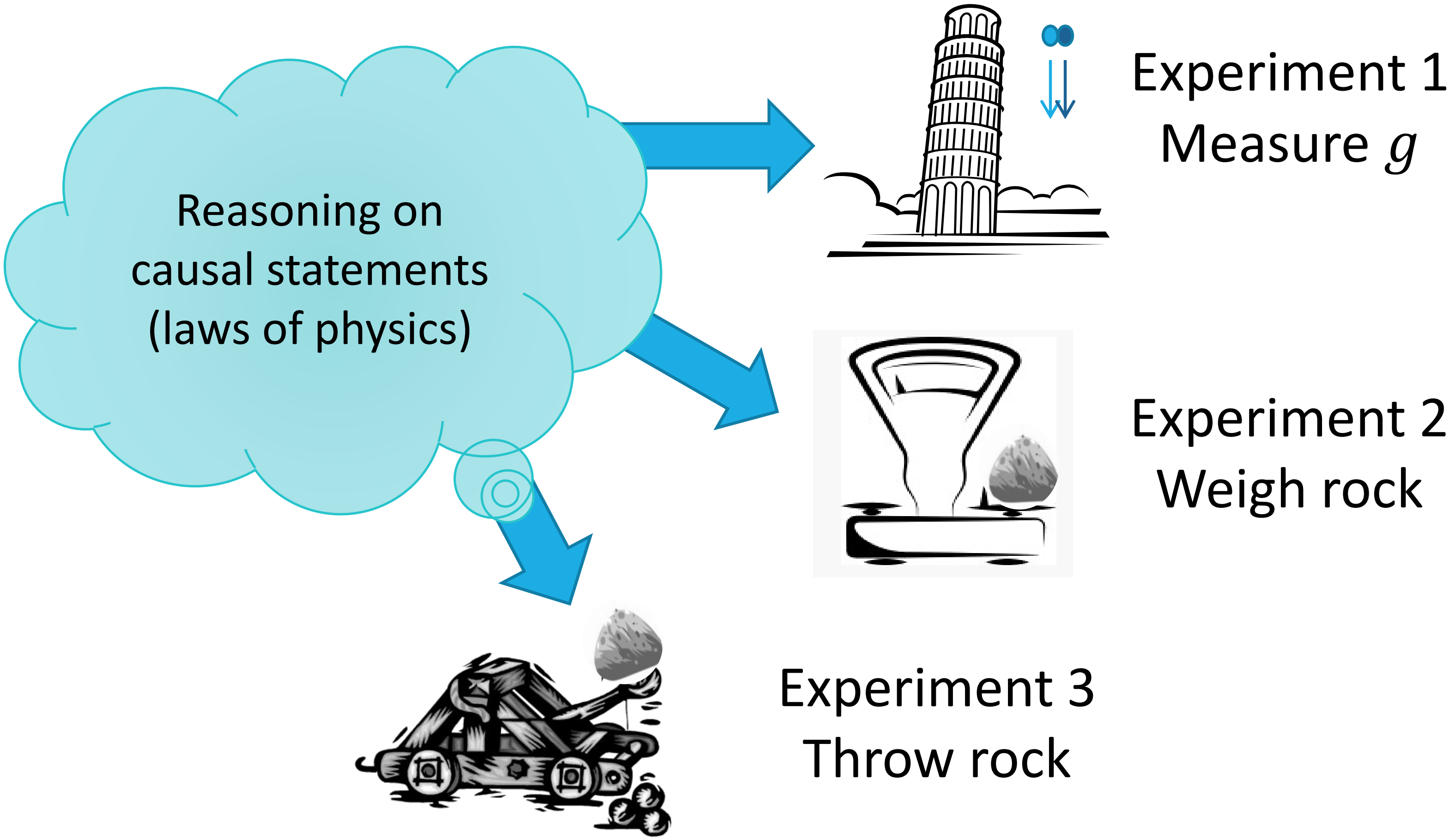
$$
\begin{aligned}
\mathrm{P}^*(\omega) = \;& \mathrm{P}(u, v) \\
& \times \mathrm{P}(x \mid u) \\
& \times \cancel{\mathrm{P}(a \mid x, v)} \quad \mathrm{P}^*(q \mid x, a) \\
& \times \mathrm{P}(b \mid x, v) \\
& \times \mathrm{P}(q \mid x, a) \\
& \times \mathrm{P}(s \mid a, q, b) \\
& \times \mathrm{P}(c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

Distribution under intervention

Many related Bayes networks are born (Pearl, 2000)

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

Transfer learning on steroids



Transportation

Pearl's "transportation" problem

- How to analytically derive conditional probabilities for experiment 3 using conditional probabilities known from experiments 1 and 2 ?
- Analogous to Bayes' rule : *do*-calculus.

"Known" versus "estimated"

- Certain conditional probabilities are known (because we have coded them)
- Certain conditional probabilities can be estimated from experiments.

  → This can be difficult (e.g. continuous or high cardinality variables.)

  → Transportation can inflate estimation errors.

4. Counterfactuals

Counterfactuals

Measuring something that did not happen

"How would have the system performed if, when the data was collected, we had used scoring model M' instead of model M?"

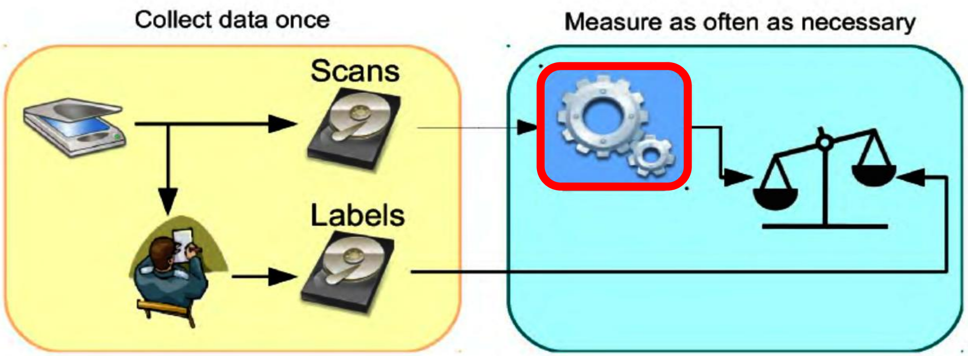
Learning procedure

- 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…

Replaying past data

Classification example

- Collect labeled data in existing setup
- Replay the past data to evaluate what the performance would have been if we had used classifier θ.

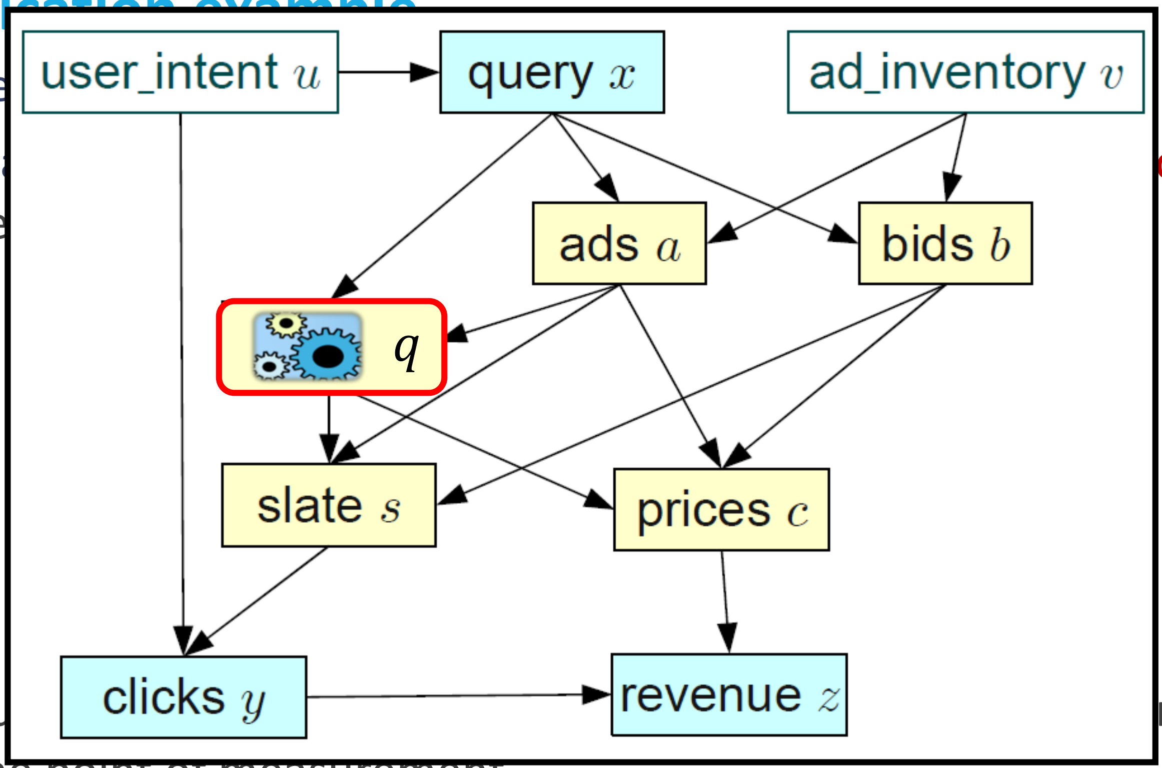


- Requires knowledge of all functions connecting the point of intervention to the point of measurement.

Replaying past data

Classifi

- Colle
- Repl d have been
  if we



- Requ ntervention
  to the point of measurement.

Importance sampling

Distribution under intervention

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

$$\mathbf{P}^{*}(q \mid x, a)$$

Importance sampling

Actual expectation

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

Counterfactual expectation*

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

* Counterfactual expectations elude the subtleties of per-item counterfactuals…

Importance sampling

Principle

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

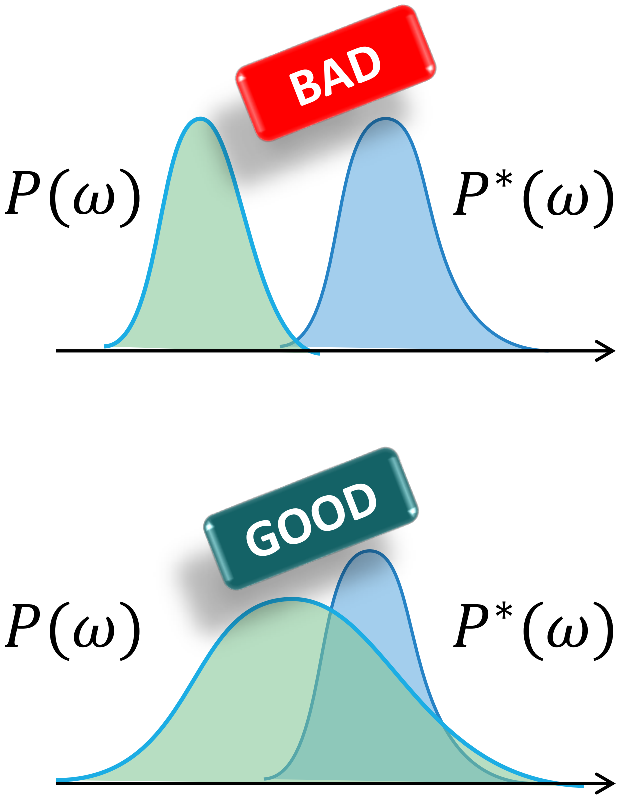
$$w(\omega_i) = \frac{P^*(\omega_i)}{P(\omega_i)} = \frac{P^*(q|x,a)}{P(q|x,a)}$$

Factors in P* not in P

Factors in P not in P*

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

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)

Well explored area

$$\Omega_R = \{\, \omega : P^*(\omega) < R\; P(\omega)\,\}$$

Easier estimate

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

$$\text{with } \bar{w}(\omega) = \begin{cases} w(\omega) & \text{if } \omega \in \Omega_R \\ 0 & \text{otherwise} \end{cases}$$

This works because $0 \le \bar{w}(\omega) \le R$.

Confidence intervals (iii)

Bounding the bias

Assuming $0 \leq \ell(\omega) \leq M$ we have

$$0 \leq Y^* - \bar{Y}^* \leq \int_{\Omega \setminus \Omega_R} \ell(\omega) P^*(\omega) \leq M \, P^*\{\Omega \setminus \Omega_R\} = M[1 - P^*(\Omega_R)]$$

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

- This is easy to estimate because $\bar{w}(\omega)$ is bounded.
- This represents the cost of insufficient exploration.
- Bonus: this remains true if $P(\omega)$ is zero in some places

Two-parts confidence interval

$$\mathbf{Y}^* - \overline{\mathbf{Y}}_\mathbf{n}^* = (\boldsymbol{Y}^* - \overline{\boldsymbol{Y}}^*) + (\overline{\mathbf{Y}}^* - \overline{\mathbf{Y}}_{\boldsymbol{n}}^*)$$

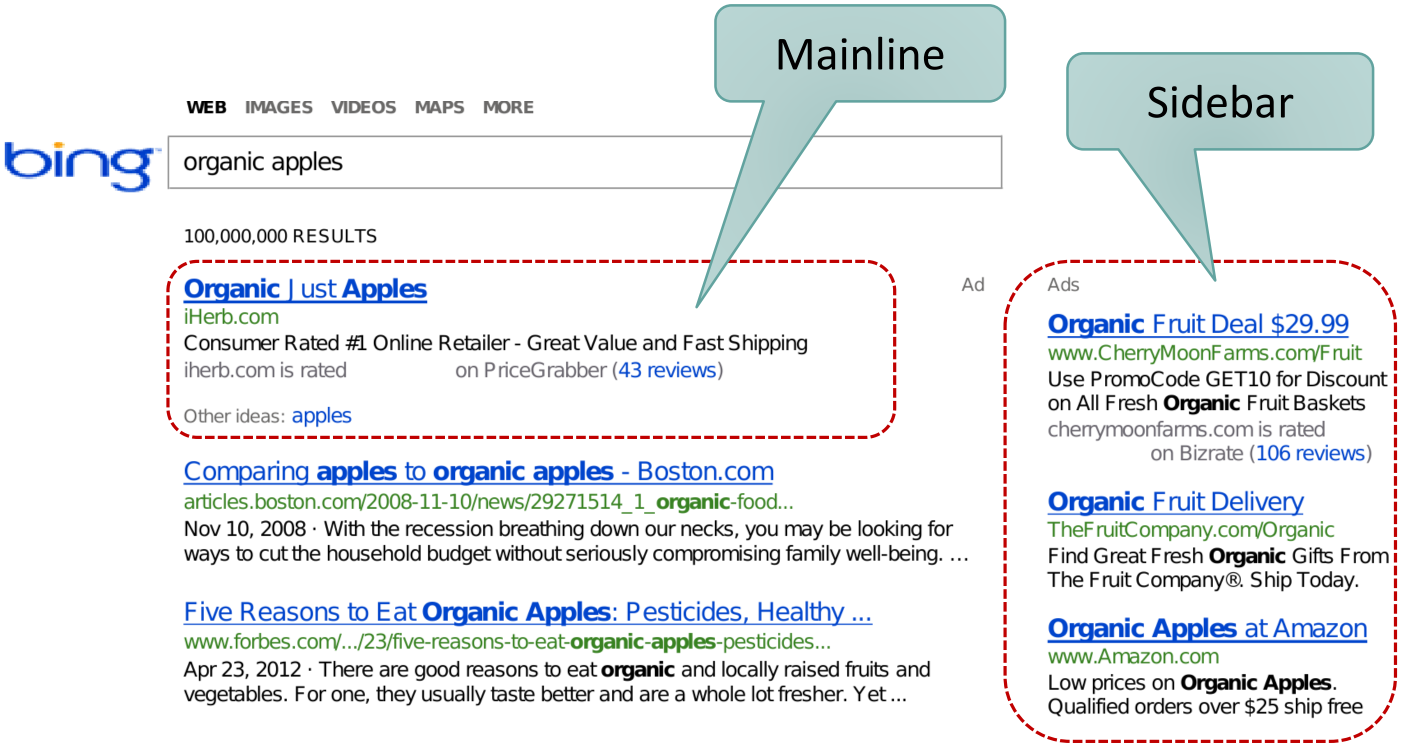
Outer confidence interval

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

Inner confidence interval

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

Illustration: Mainline ads



Playing with mainline reserves

Mainline reserves (MLRs)

- Rank score thresholds that control whether ads are displayed above the search results.

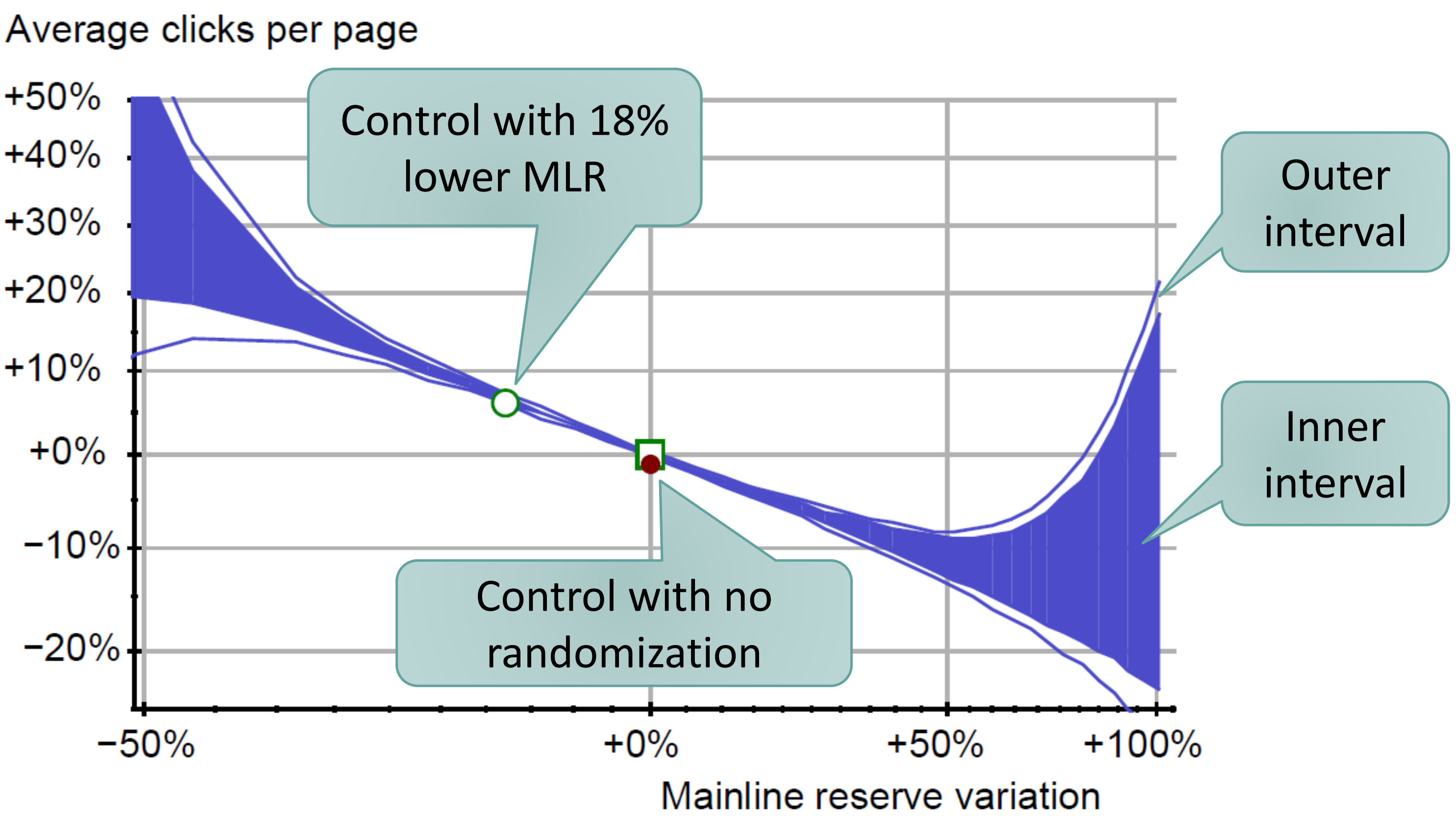
Data collection bucket

- Random log-normal multiplier applied to MLRs.
- 22M auctions over five weeks (summer 2010)

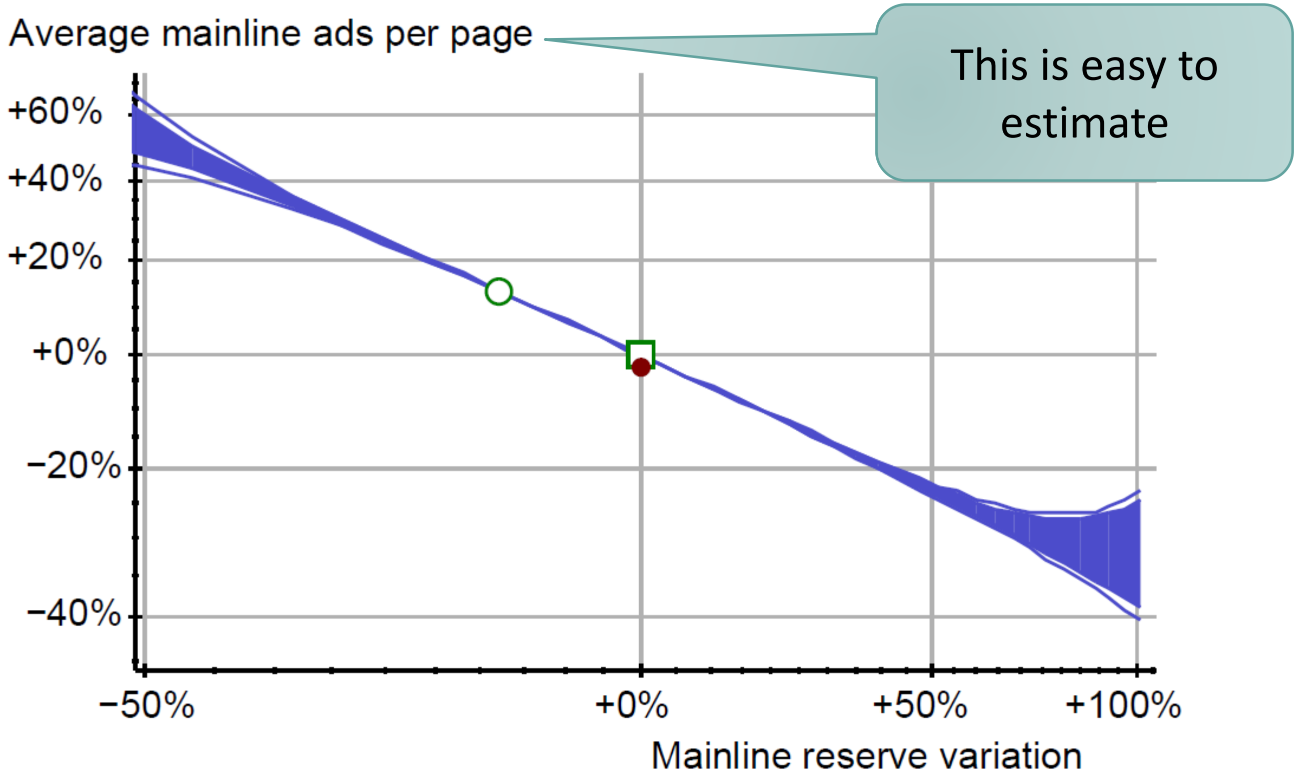
Control buckets

- Same setup with 18% lower mainline reserves
- Same setup without randomization

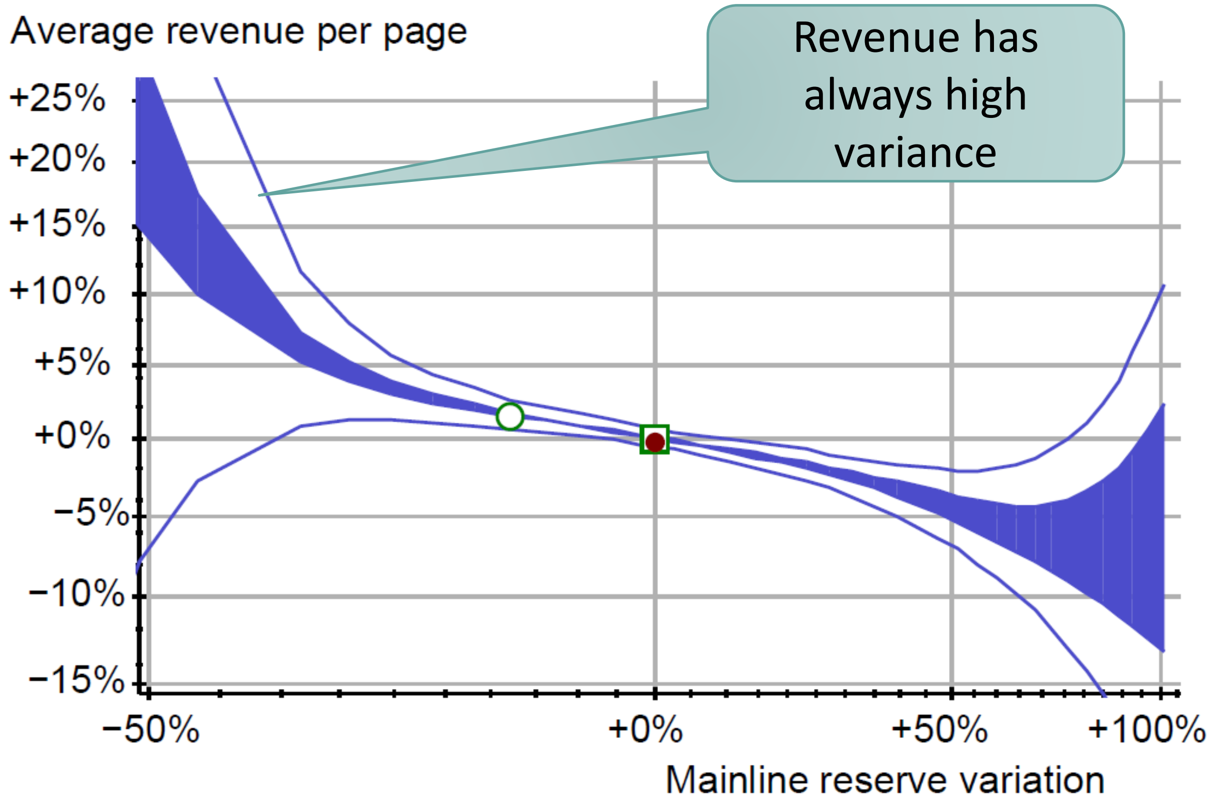
Playing with mainline reserves



Playing with mainline reserves



Playing with mainline reserves



More uses for the same data

Examples

Estimates for different randomization variance

→ Good to determine how much to explore.

Query-dependent reserves

→ Just another counterfactual distribution!

This is the big advantage

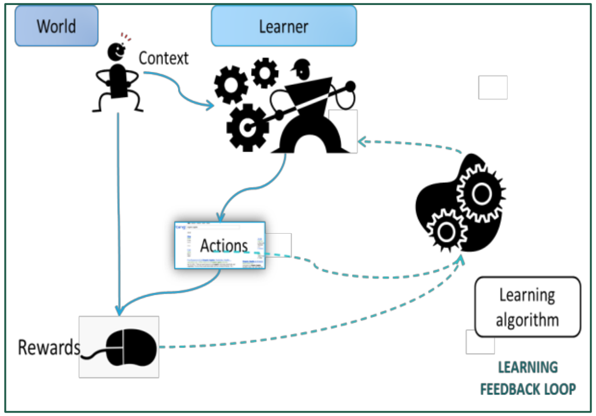
- Collect data first, choose questions later.
- Randomizing more stuff increases opportunities.
- New challenge: making sure that do not leave information on the table.

5. Structure

Contextual bandits

Framework

- World select context x
- Learner chooses discrete action
  $a = \pi(x) \in \{1 \dots K\}$
- World announces reward $r(x, a)$



Results

- Randomized data collection (i.e., exploration) enables offline unbiased evaluation of an alternate policy π^* by means of importance sampling.
- Solid analysis of the explore/exploit trade-off, that is, how much exploration is needed at each instant.

(Langford et al., 2008) (Li et. al., 2010, 2011)

Structure

Actions have structure

- What we learn by showing a particular ad for a particular query
  tells us about showing similar ads for similar queries.

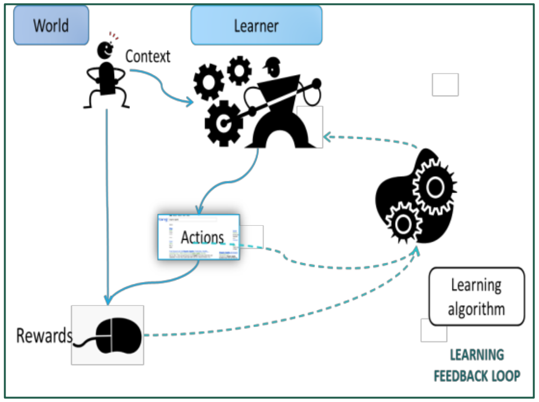
Policies have structure

- One action is a set of ads displayed on a page.
  But computationally feasible policies score each ad individually.

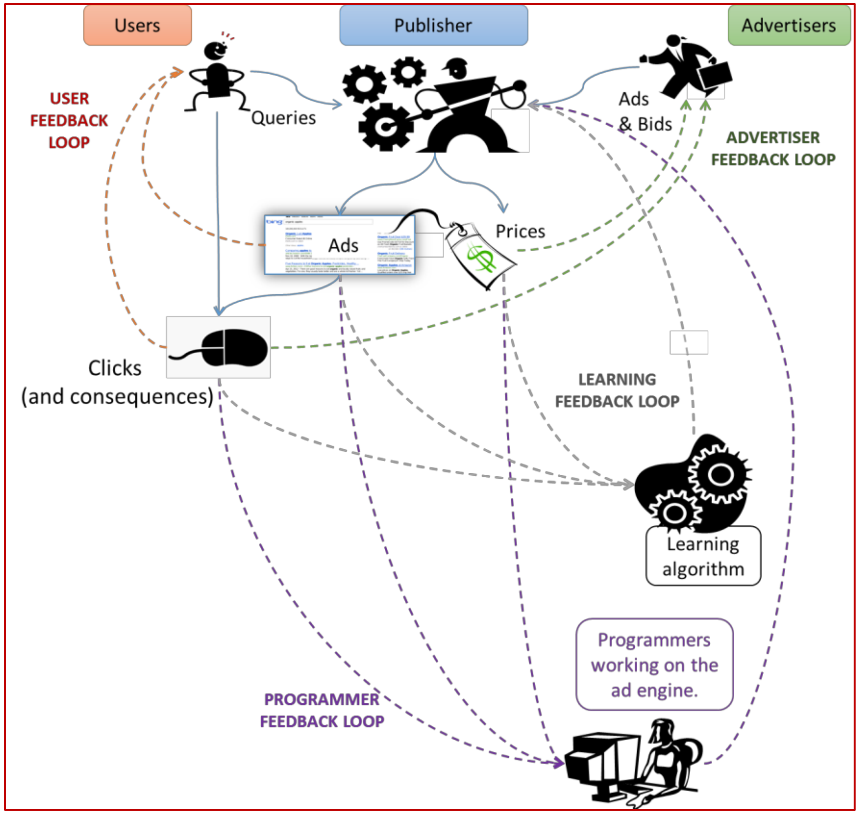
Rewards have structure

- Actions are set of ads with associated click prices.
  Chosen ads impact users, chosen prices impact advertisers.

The causal graph has structure



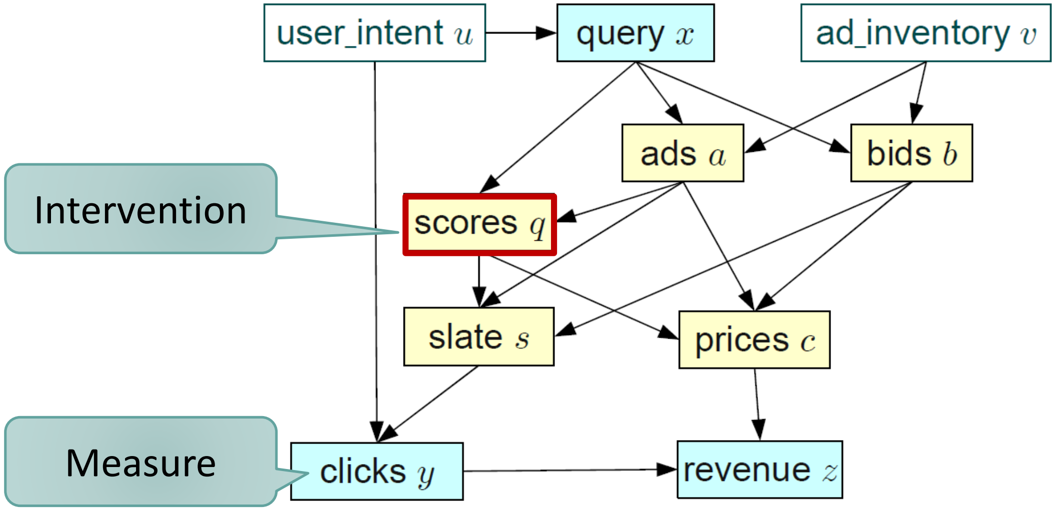
≠



Using the causal structure (1)

Improved confidence intervals

- Example: users click without knowing the ad scores or the click prices.
- Technique: “shifting” the reweighting variables



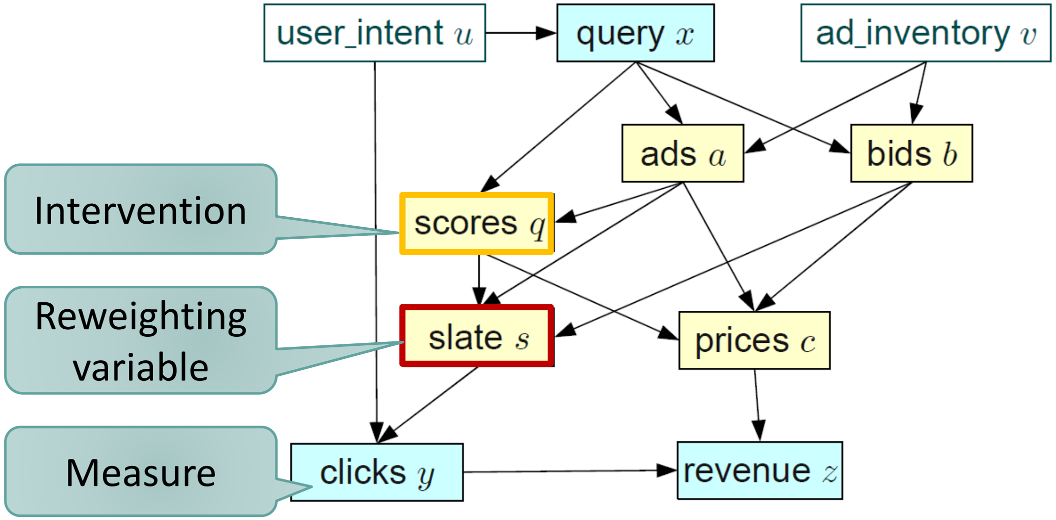
Standard weights

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

Using the causal structure (1)

Improved confidence intervals

- Example: users click without knowing the ad scores or the click prices.
- Technique: “shifting” the reweighting variables



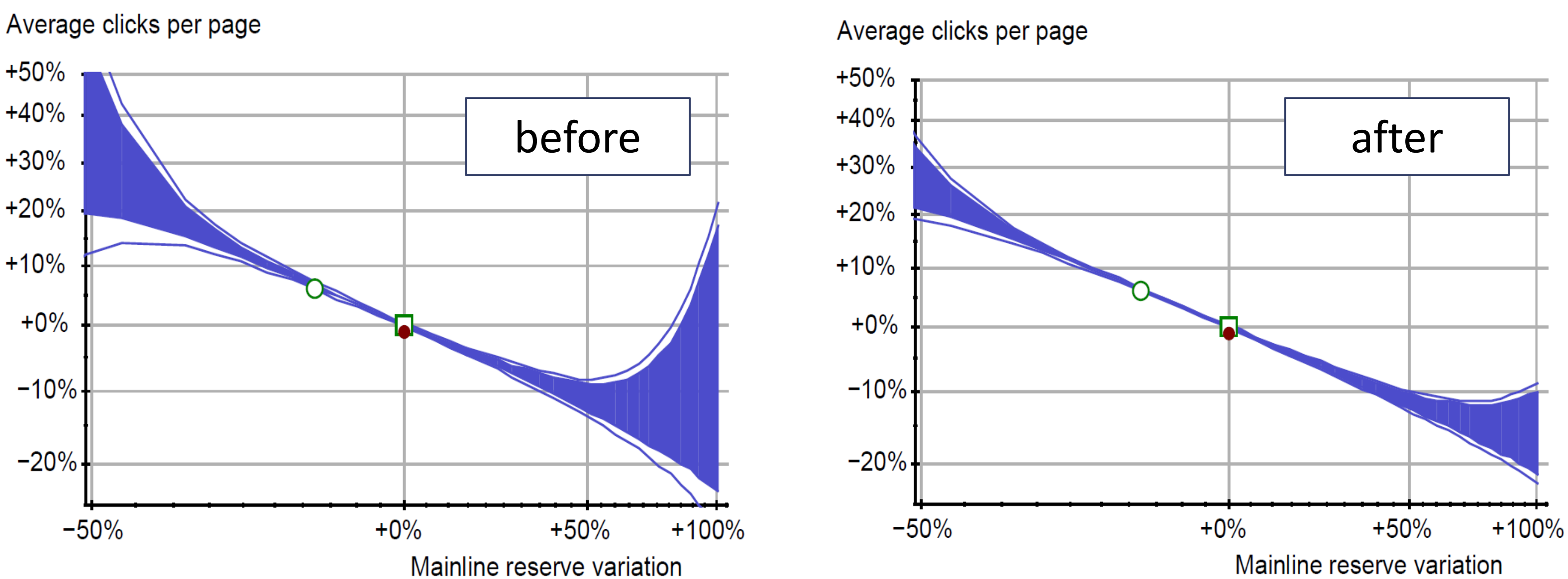
Standard weights

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

Shifted weights

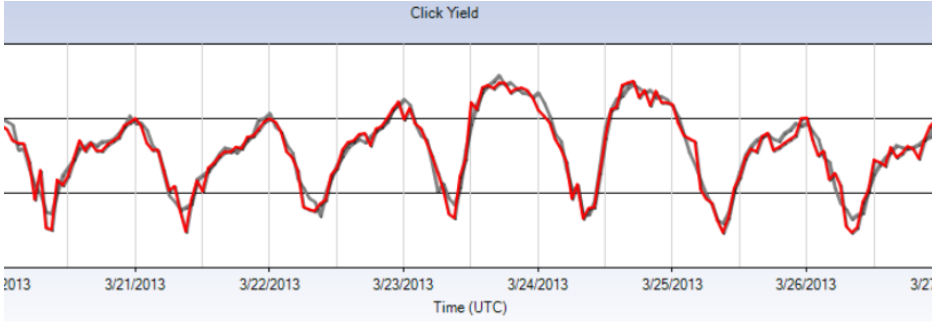
$$w(\omega_i) = \frac{P^*(\omega_i)}{P(\omega_i)} = \frac{P^*(s|x,a,b)}{P(s|x,a,b)}$$

Illustration



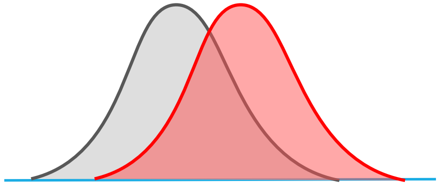
Variance reduction for A/B testing

Hourly average click yield for treatment and control



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

Daily effects increases the variance of both treatment and control.



Daily effects affect treatment and control in similar ways!
Can we subtract them?

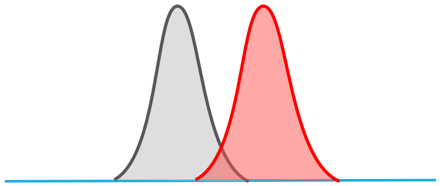
Variance reduction for A/B testing

- Treatment estimate $Y^* \approx \hat{Y}^* = \frac{1}{|T|} \sum_{i \in T} y_i$
- Control estimate $Y \approx \hat{Y} = \frac{1}{|C|} \sum_{i \in C} y_i$

- Predictor $\zeta(x)$ estimates y on the basis of only the query time x.
- Then $Y^* - Y = (Y^* - \mathbb{E}[\zeta(x)]) - (Y - \mathbb{E}[\zeta(x)])$

$$\approx \frac{1}{|T|} \sum_{i \in T} \left(y_i - \zeta(x_i) \right) - \frac{1}{|C|} \sum_{i \in C} \left(y_i - \zeta(x_i) \right)$$

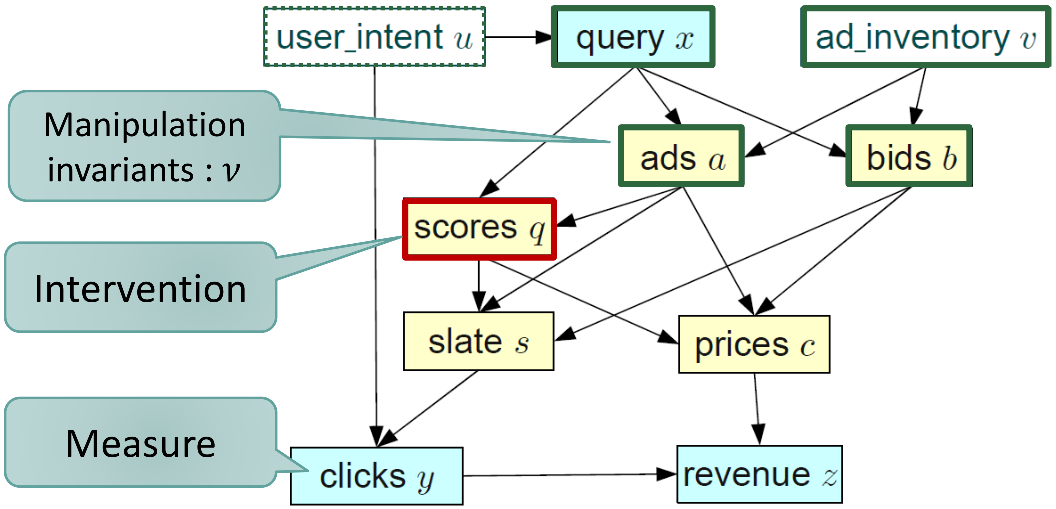
This is true regardless of the predictor quality.
But if it is any good, $\mathrm{var}[Y - \zeta(X)] < \mathrm{var}[Y]$, and



Using the causal structure (2)

Manipulation invariant variables

- Example: changes in query time introduces unwanted variance .
- Technique: leveraging an invariant predictor



Manipulation invariant variables

The probability distribution of ν is not affected by the manipulation of the causal model.

Estimating differences

Comparing two potential interventions

Is scoring model M_1 better than M_2 ?

$$\Delta = \boxed{\text{Click-thru-rate if we had used model } M_1} - \boxed{\text{Click-thru-rate if we had used model } M_2}$$

Improved confidence via variance reduction

- Example: since seasonal variations affect both models nearly identically, the variance resulting from these variations cancels in the difference.

Estimating differences

Which scoring model works best?

- Comparing expectations under counterfactual distributions $P^+(\omega)$ and $P^*(\omega)$.

$$Y^+ - Y^* = \int_\omega [\ell(\omega) - \zeta(\nu)]\, \Delta w(\omega)\ P(\omega)$$
$$\approx \frac{1}{n}\sum_{i=1}^{n} [\ell(\omega_i) - \zeta(\nu_i)]\, \Delta w(\omega_i)$$

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

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

6. Learning

Estimating derivatives

Infinitesimal interventions

$$\frac{\partial CTR}{\partial \theta} = \left(\boxed{\text{Click rate if we had used } M(\theta + d\theta)} - \boxed{\text{Click rate if we had used model } M(\theta)} \right) \Big/ \mathrm{d}\theta$$

- Related to “policy gradient” in RL.
- Optimization algorithms learn model parameters θ.

Estimating derivatives

Counterfactual distribution $P^\theta(\omega)$

$$\frac{\partial Y^\theta}{\partial \theta} = \int_\omega [\ell(\omega) - \zeta(\nu)]\, w'_\theta(\omega) P(\omega) \approx \frac{1}{n} \sum_{i=1}^{n} [\ell(\omega_i) - \zeta(\nu_i)]\, w'_\theta(\omega_i)$$

with $w_\theta(\omega) = \frac{P^\theta(\omega)}{P(\omega)}$ and $w'_\theta(\omega) = \frac{\partial w_\theta(\omega)}{\partial \theta} = \frac{P^\theta(\omega)}{P(\omega)} \frac{\partial \log P^\theta(\omega)}{\partial \theta}$

This ratio can be large but …

Policy gradient

Infinitesimal manipulation

- Collect data pertaining to distribution $P^{\theta}(\omega)$
- Estimate performance metrics for $P^{\theta+d\theta}(\omega)$

$$Y^{\theta+d\theta} = Y^{\theta} + \left.\frac{\partial Y^{\theta}}{\partial \theta}\right|_{\theta} \times d\theta$$

$$\text{with} \quad \left.\frac{\partial Y^{\theta}}{\partial \theta}\right|_{\theta} \approx \frac{1}{n}\sum_{i=1}^{n}[\ell(\omega_i) - \zeta(\nu_i)] \cancel{\frac{P^{\theta}(\omega)}{P^{\theta}(\omega)}} \frac{\partial \log P^{\theta}(\omega)}{\partial \theta}$$

… it goes away for small interventions

(Williams, 1992)

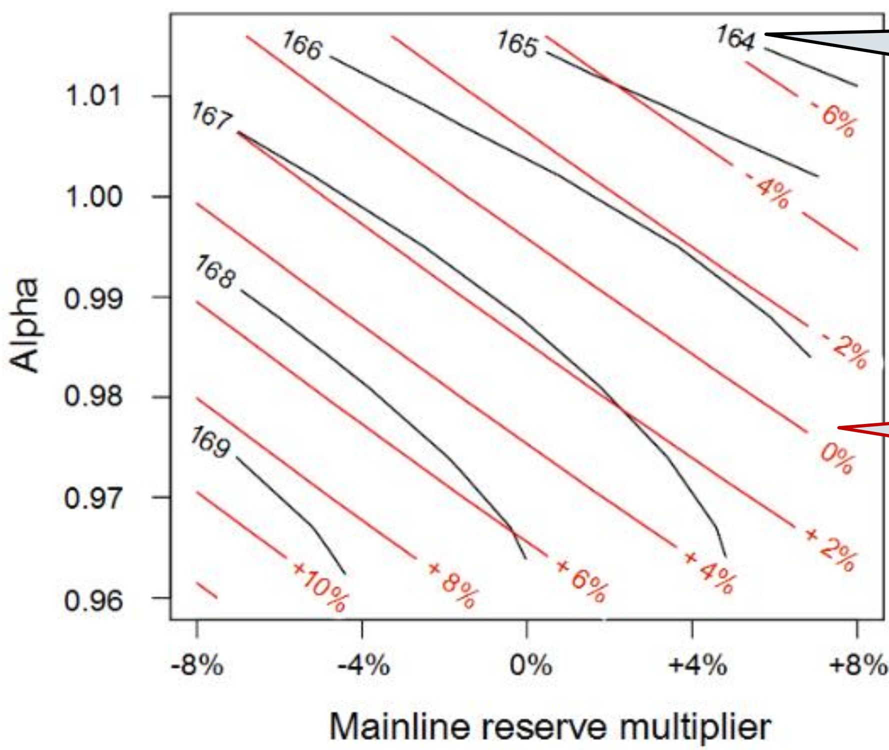
Derivatives and optimization

Example: tuning squashing exponents and reserves

- Ads ranked by decreasing $bid \times pClick^{\alpha}$
- Lahaie and McAfee (2011) show that using $\alpha < 1$ is good when click probability estimation gets less accurate.
- Different α_k and reserves ρ_k for each query cluster k.

Derivatives and optimization

Level curves for one particular query cluster



Estimated advertiser value (arbitrary units)

Variation of the average number of mainline ads.

Learning as counterfactual optimization

- Does it generalize?

  Yes, we can obtain uniform confidence intervals.
- Sequential design?

  Thompson sampling comes naturally in this context.
- Metering exploration wisely?

  Inner confidence interval tells how much exploration we need to answer a counterfactual question.

  But this does not tell which questions we should ask.

  This was not a problem in practice…

7. Equilibrium

Revisiting the feedback loops

Tracking the equilibrium

If we increase the ad relevance thresholds :

- We show less ads and lose revenue in the short time.
- Users see more relevant ads, are more likely to click on ads in the future, possibly making up for the lost revenue [eventually].
- Advertisers will [eventually] update their bids.
  It could go both ways because they receive less clicks
  from more engaged users…

Counterfactual equilibrium

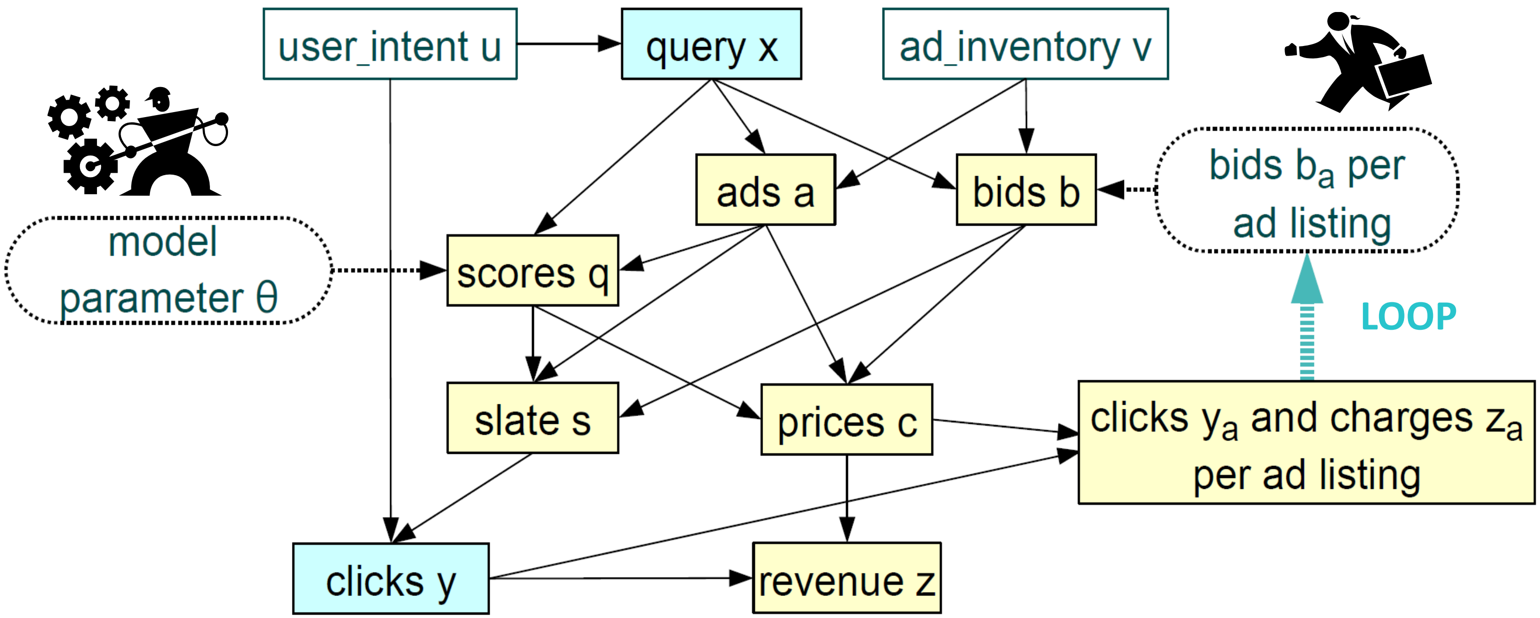
Counterfactual question

> *"What would have been the system performance metrics if we had applied an infinitesimal change to the parameter θ of the scoring model long enough to reach the equilibrium during the data collection period?"*

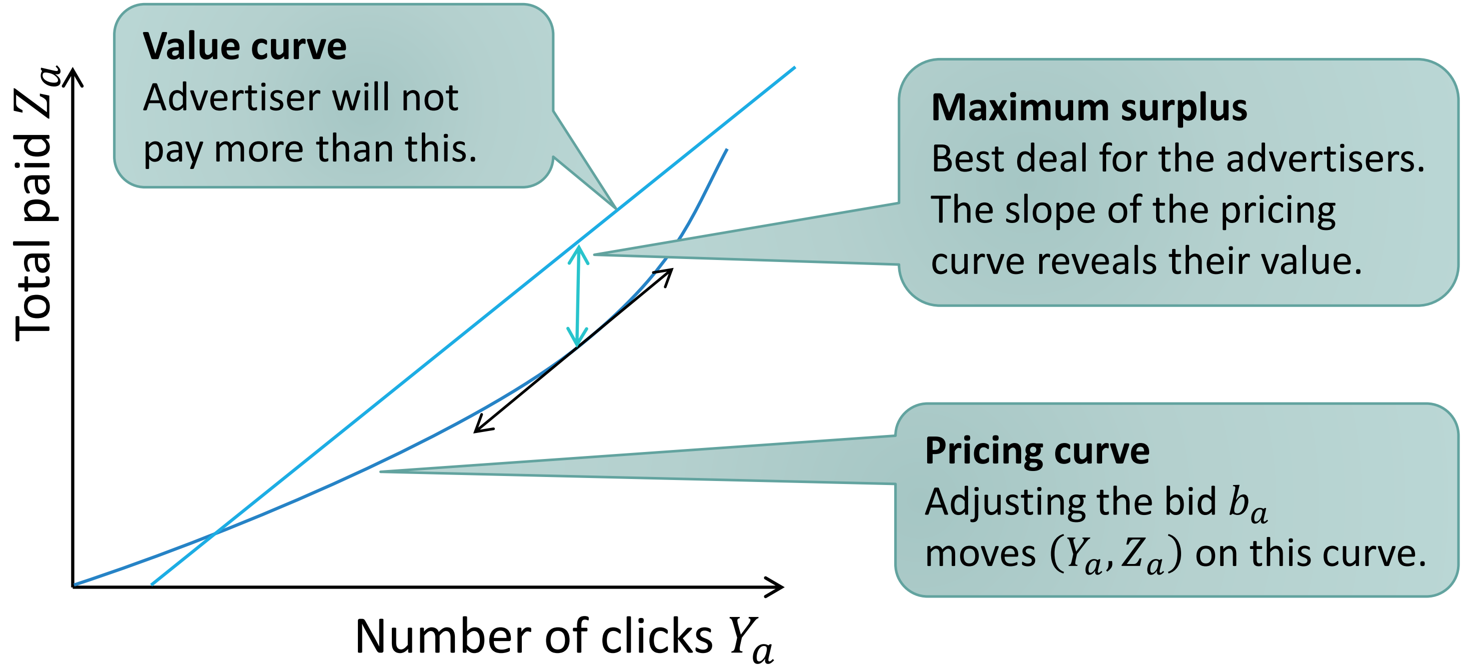
We can answer using *quasi-static analysis*.

Thanks to the work of many physicists…

Advertiser feedback loop



Rational advertiser



Rational advertisers keep $\mathrm{V_a} = \frac{\partial Z_a}{\partial Y_a} = \frac{\partial Z_a}{\partial b_a} \Big/ \frac{\partial Y_a}{\partial b_a}$ constant!

(Athey & Nekipelov, 2010)

Estimating values

When the system reaches equilibrium, we can compute

$$V_a = \frac{\partial Z_a}{\partial b_a} \Big/ \frac{\partial Y_a}{\partial b_a} = \frac{\partial E_{\boldsymbol{b},\theta}(z_a)}{\partial b_a} \Big/ \frac{\partial E_{\boldsymbol{b},\theta}(y_a)}{\partial b_a}$$

Counterfactual derivatives

- Complication: we cannot randomize the bids. However, since ads are ranked by *bids×scores*, we can interpret a random score multiplier as a random bid multiplier (need to reprice.)

Feedback loop equilibrium

Derivative of surplus vector $\Phi = \left[\dots \frac{\partial Z_a}{\partial b_a} - V_a \frac{\partial Y_a}{\partial b_a} \dots \right] = 0$.

$$\mathrm{d}\Phi = \frac{\partial \Phi}{\partial \theta} \mathrm{d}\theta + \sum_a \frac{\partial \Phi}{\partial b_a} \mathrm{d}b_a = 0$$

Solving the linear system yields $\frac{\mathrm{d}b_a}{\mathrm{d}\theta}$.

Equilibrium condition

Then we answer the counterfactual question

$$\mathrm{d}Y = \left(\frac{\partial Y}{\partial \theta} + \sum_a \frac{\partial Y}{\partial b_a} \frac{\mathrm{d}b_a}{\mathrm{d}\theta} \right) \mathrm{d}\theta$$

Multiple feedback loops

Same procedure

1. Write total derivatives.
2. Solve the linear system formed
   by all the equilibrium conditions.
3. Substitute into the total derivative
   of the counterfactual expectation of interest.

Conclusion

Main messages

- Relation between explore-exploit and correlation-causation.
- The causation framework provides a rich and modular framework for engineering of web-scale interactive learning systems
- The differential equilibrium analysis methods of physics apply.
- Cybernetics!

Tech report available at http://leon.bottou.org/papers.