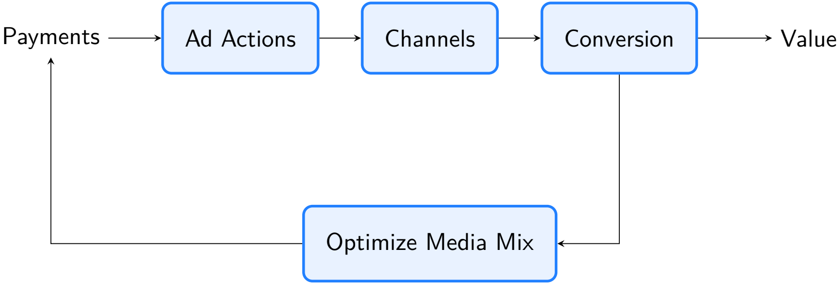
An Axiomatic Framework for Multi-channel Attribution in Online Advertising

Garud Iyengar
Columbia University

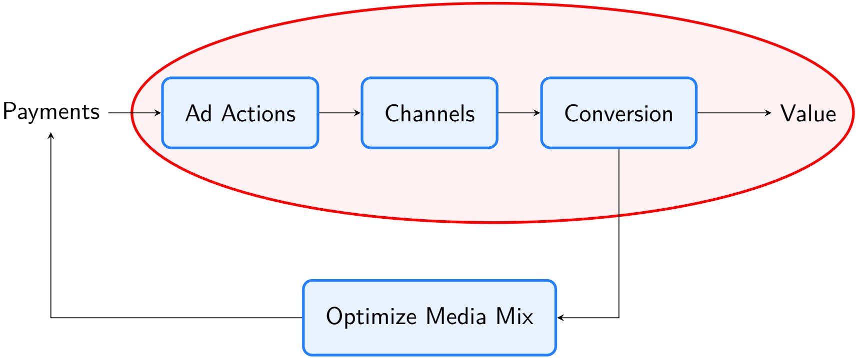
Workshop on Marketplace Innovation 2017

Joint work with Antoine Desir, Vineet Goyal and Omar Besbes
Partially support by a grant from Adobe

Optimization Cycle in Online Advertising



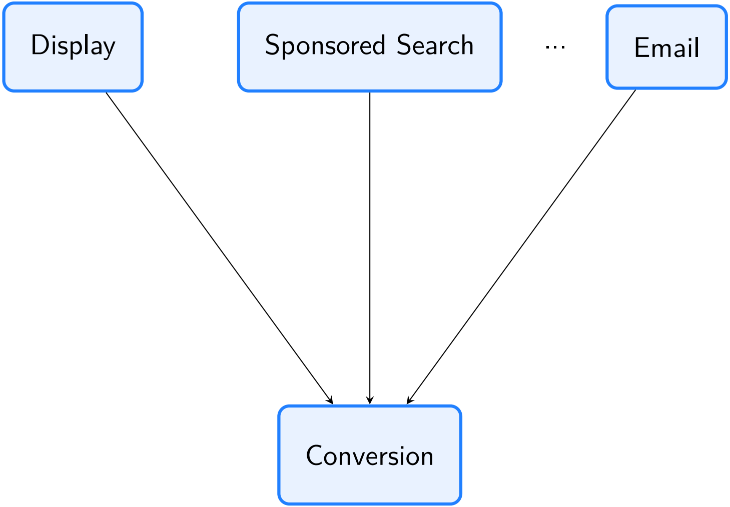
Optimization Cycle in Online Advertising



A lot of academic effort and progress on media mix optimization . . .

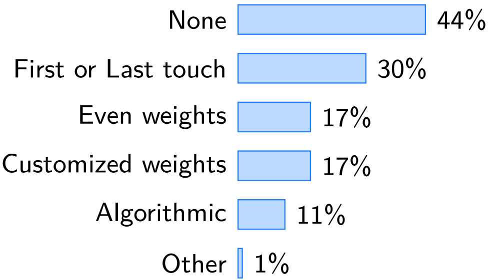
relatively less progress on the measurement and attribution steps

Multi-Channel Attribution in Online Advertising



Question: How to give credit to all channels contributing to a conversion?

Common current attribution methods



Source: *Beyond Last Touch: Understanding Campaign Effectiveness*, 2013 Quantcast

Goals for this work

- Framework for attributing credit to channels
- Evaluate heuristic procedures
- Address limitations of current heuristics and propose prescriptions

Related work

Heuristics for multi-touch attribution

- Shapley value: [Shao and Li, 2011], [Dalessandro et al., 2012], [Berman, 2013], [Li and Kannan, 2014]
- Incremental impact: [Abhishek et al., 2012], [Anderl et al., 2014]

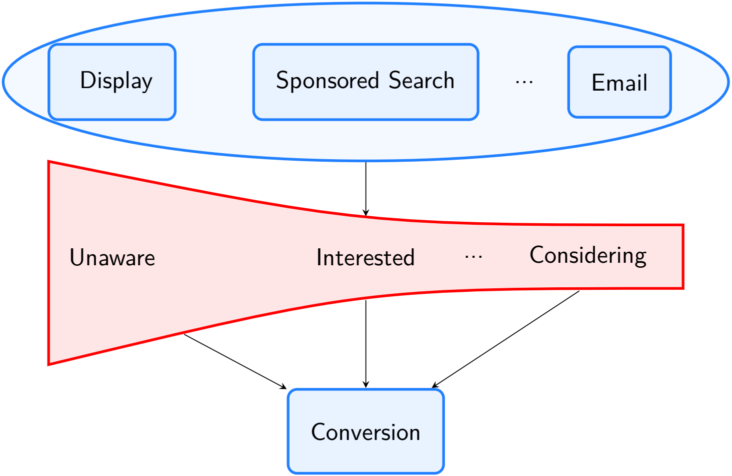
Empirical work

- [Kireyev et al., 2013], [Lewis and Rao, 2014]

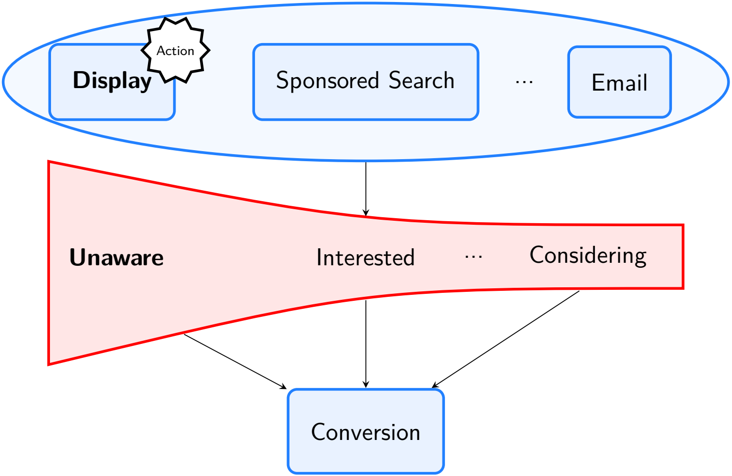
Other related work

- [Jordan et al., 2011], [Archak et al., 2010]

Model elements: channels and customer states

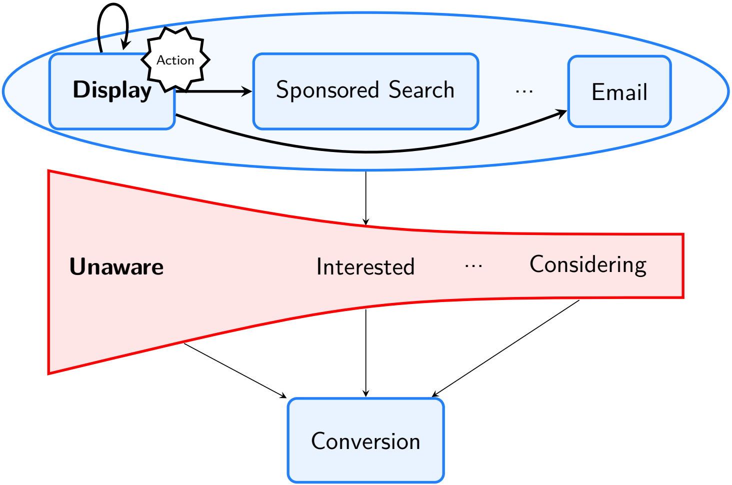


Model elements: channels and customer states



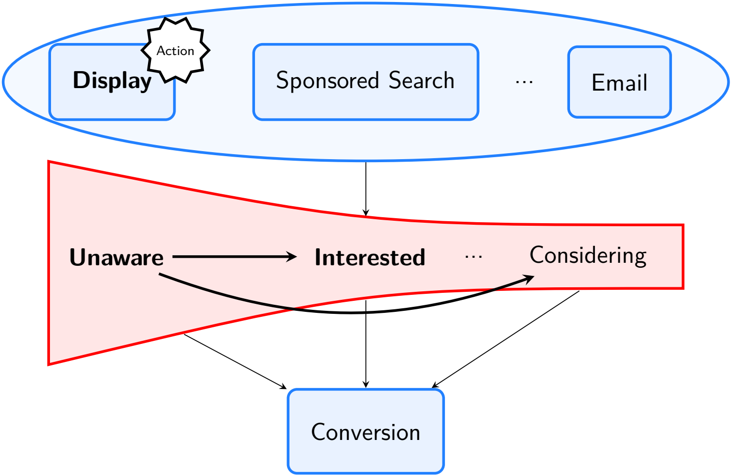
Two possible effects of an action:

Model elements: channels and customer states



Two possible effects of an action: Network effect

Model elements: channels and customer states



Two possible effects of an action: Network effect and Funnel effect

Approach

"Observational" inference

- Each channel c shows the ad to β_c fraction of customers
- Observe one-step browsing behavior

Attribution

- Payment per impression
- Payment independent of the customer state or browsing behavior

Model

Markov chain consisting two state variables $\sigma = (c, s)$

- $c =$ identity of the channel, e.g. particular website
- $s =$ latent state of the visitor

Will ignore latent state in the talk. All results go through!

Model

Markov chain consisting two state variables $\sigma = (c, s)$

- $c =$ identity of the channel, e.g. particular website
- $s =$ latent state of the visitor

Will ignore latent state in the talk. All results go through!

Transition probability

- $P^{(1)}_{c,c'} =$ probability $c \to c'$ when a visitor
  exposed to an ad on channel c
- $P^{(0)}_{c,c'} =$ probability $c \to c'$ when a visitor
  not exposed to an ad on channel c
- Transition when β fraction see an ad

$$P^{\beta} = \mathbf{diag}(\beta)P^{(1)} + \mathbf{diag}(\mathbf{1} - \beta)P^{(0)}$$

Model (contd)

- Two absorbing states
    - c^* = conversion. Conversion only on advertising.
    - q = quit
- π_c^* = conversion probability from channel c

Model (contd)

- Two absorbing states
  - $c^* =$ conversion. Conversion only on advertising.
  - $q =$ quit
- $\pi_c^* =$ conversion probability from channel c
- $\lambda_c =$ external traffic to channel c and $\mu_c^\beta =$ effective traffic at c

$$\mu^\beta = \lambda + (P^\beta)^\top \mu^\beta$$

- $h_c^\beta =$ conversion probability from c

$$h^\beta = P^\beta h^\beta + \pi^*$$

Payment per **impression**

Channel c is paid z_c per impression

- Payment **not** a function of the latent state
- Payment **not** a function of the downstream browsing

Payment per impression

Channel c is paid z_c per impression

- Payment **not** a function of the latent state
- Payment **not** a function of the downstream browsing

Desirable properties

- **Channel rationality**: $z \geq 0$
- **Advertiser rationality**: global budget balance

$$\underbrace{\lambda^\top h^\beta}_{\text{Total value generated}} \geq \underbrace{(\mu^\beta)^\top \mathbf{diag}(\beta) z}_{\text{Total payments}}$$

- **Fairness**: hard to quantity!

Last touch

Pay each channel one unit for each conversion

Not quite allowed in our formulation but admits equivalent representation through payment per impression

Trivially satisfies global budget balance. But, very unfair!

Last touch

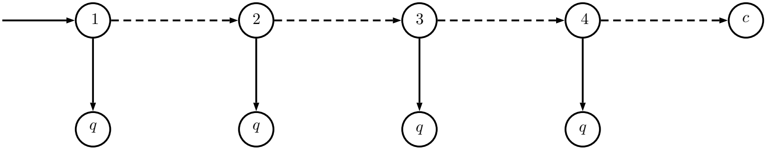
Pay each channel one unit for each conversion

Not quite allowed in our formulation but admits equivalent representation through payment per impression

Trivially satisfies global budget balance. But, very unfair!

Example

$$P^{(1)}_{i,i+1} = 1 \quad P^{(0)} = \mathbf{0} \quad \lambda_1 = 1,\ \lambda_i = 0,\ \forall i \neq 1 \quad \beta_i \equiv 1$$



All value goes to channel 4 although all traffic brought in by channel 1!

Last touch

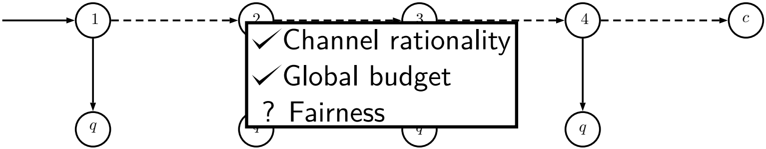
Pay each channel one unit for each conversion

Not quite allowed in our formulation but admits equivalent representation through payment per impression

Trivially satisfies global budget balance. But, very unfair!

Example

$$P^{(1)}_{i,i+1} = 1 \quad P^{(0)} = \mathbf{0} \quad \lambda_1 = 1,\ \lambda_i = 0,\ \forall i \neq 1 \quad \beta_i \equiv 1$$



All value goes to channel 4 although all traffic brought in by channel 1!

Incremental value heuristic (IVH)

Incremental value of the advertising action

- Advertise: $P^{(1)} h^{\beta} + \pi^*$
- Do not advertise: $P^{(0)} h^{\beta}$
- Incremental value: $\underbrace{(P^{(1)} - P^{(0)})}_{:= \Delta} h^{\beta} + \pi^*$

Incremental value heuristic (IVH)

Incremental value of the advertising action

- Advertise: $P^{(1)}h^{\beta} + \pi^*$
- Do not advertise: $P^{(0)}h^{\beta}$
- Incremental value: $\underbrace{\left(P^{(1)} - P^{(0)}\right)}_{:= \Delta} h^{\beta} + \pi^*$

Incremental value heuristic: $z = \Delta h^{\beta} + \pi^*$. Pays for both

- immediate conversion: π^*
- network effect: Δh^{β}

Incremental value heuristic (IVH)

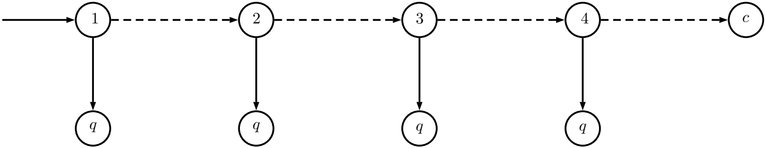
Incremental value of the advertising action

- Advertise: $P^{(1)}h^\beta + \pi^*$
- Do not advertise: $P^{(0)}h^\beta$
- Incremental value: $\underbrace{\left(P^{(1)} - P^{(0)}\right)}_{:=\,\Delta} h^\beta + \pi^*$

Incremental value heuristic: $z = \Delta h^\beta + \pi^*$. Pays for both

- immediate conversion: π^*
- network effect: Δh^β

Example: $z_c \equiv 1$ fair! But $\sum_c \mu_c^\beta \beta_c z_c = n \gg \sum_c \lambda_c h_c^\beta = 1$!



Incremental value heuristic (IVH)

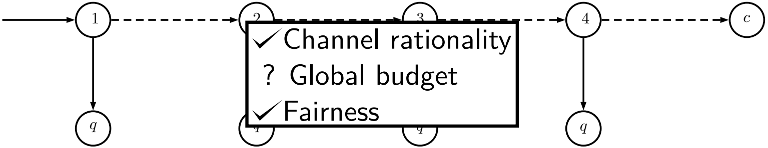
Incremental value of the advertising action

- Advertise: $P^{(1)} h^{\beta} + \pi^*$
- Do not advertise: $P^{(0)} h^{\beta}$
- Incremental value: $\underbrace{\left(P^{(1)} - P^{(0)}\right)}_{:=\,\Delta} h^{\beta} + \pi^*$

Incremental value heuristic: $z = \Delta h^{\beta} + \pi^*$. Pays for both

- immediate conversion: π^*
- network effect: Δh^{β}

Example: $z_c \equiv 1$ fair! But $\sum_c \mu_c^{\beta} \beta_c z_c = n \gg \sum_c \lambda_c h_c^{\beta} = 1!$



✓ Channel rationality
? Global budget
✓ Fairness

IVH: generically overbudget

Proposition. Let $\delta = \mu^T \text{diag}(\beta) \Delta h(\beta)$. Then

$$\lambda^\top h^\beta - (\mu^\beta)^\top \mathbf{diag}(\beta) z = -\delta$$

$\delta > 0 \approx$ when advertising generates downstream indirect conversions

IVH: generically overbudget

Proposition. Let $\delta = \mu^T \text{diag}(\beta) \Delta h(\beta)$. Then

$$\lambda^\top h^\beta - (\mu^\beta)^\top \mathbf{diag}(\beta) z = -\delta$$

$\delta > 0 \approx$ when advertising generates downstream indirect conversions

Over payment: Every conversion gets counted several times

- pay for both for direct conversion, and
- eventual conversion

IVH with past payments

Need to account for past payments!

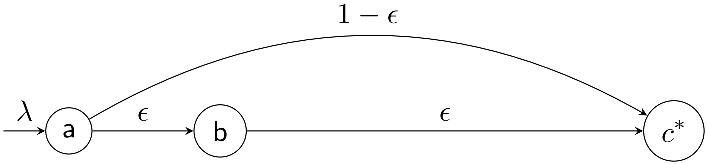
Proposal: $z_c = (\Delta h)_c + \pi_c^* - \mathbb{E}[\text{past payments}|c]$

IVH with past payments

Need to account for past payments!

Proposal: $z_c = (\Delta h)_c + \pi_c^* - \mathbb{E}[\text{past payments}|c]$

Example



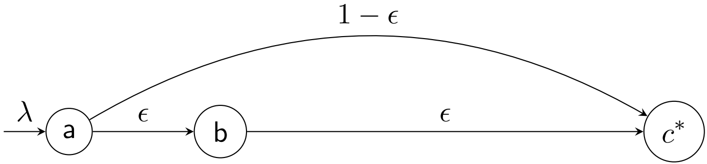
- Channel b: past payment $= z_a$ and $(\Delta h)_b + \pi_b^* = \epsilon$. Thus, $z_a \leq \epsilon$.

IVH with past payments

Need to account for past payments!

Proposal: $z_c = (\Delta h)_c + \pi_c^* - \mathbb{E}[\text{past payments}|c]$

Example



- Channel b: past payment $= z_a$ and $(\Delta h)_b + \pi_b^* = \epsilon$. Thus, $z_a \leq \epsilon$.
- Channel a: cannot be compensated for the $1-\epsilon$ fraction it converts!

Outcome based cost accounting

Requirements

- discriminate between outcomes when allocating payments
- local budget balance must ensure global budget balance

Outcome based cost accounting

Requirements

- discriminate between outcomes when allocating payments
- local budget balance must ensure global budget balance

Cost per visitor at channel c: $p_c + \beta_c z_c$

- past payment p_c
- payment per impression z_c

Outcome based cost accounting

Requirements

- discriminate between outcomes when allocating payments
- local budget balance must ensure global budget balance

Cost per visitor at channel c: $p_c + \beta_c z_c$

- past payment p_c
- payment per impression z_c

Allocate cost $x_{a,b}$ along the arc (a, b)

- Axiom 1: Cost proportional to outcome value: $x_{a,b} \propto h_b^{\beta}$

Outcome based cost accounting

Requirements

- discriminate between outcomes when allocating payments
- local budget balance must ensure global budget balance

Cost per visitor at channel c: $p_c + \beta_c z_c$

- past payment p_c
- payment per impression z_c

Allocate cost $x_{a,b}$ along the arc (a, b)

- Axiom 1: Cost proportional to outcome value: $x_{a,b} \propto h_b^\beta$
- Axiom 2: Forward payment conservation: $\mathbb{E}[x_{a,b}] = p_a + \beta_a z_a$

Outcome based cost accounting

Requirements

- discriminate between outcomes when allocating payments
- local budget balance must ensure global budget balance

Cost per visitor at channel c: $p_c + \beta_c z_c$

- past payment p_c
- payment per impression z_c

Allocate cost $x_{a,b}$ along the arc (a, b)

- Axiom 1: Cost proportional to outcome value: $x_{a,b} \propto h_b^\beta$
- Axiom 2: Forward payment conservation: $\mathbb{E}[x_{a,b}] = p_a + \beta_a z_a$
- Axiom 3: Backward payment conservation: $p_c = \sum_b \mathbb{P}(b \mid c) x_{b,c}$

Outcome based accounting (contd)

Axioms 1+2

$$x_{a,b} = (p_a + \beta_a z_a) \cdot \frac{h_b^\beta}{\sum_c P_{a,c}^\beta h_c^\beta + \pi_a^*} = (p_a + \beta_a z_a) \cdot \frac{h_b^\beta}{h_a^\beta}$$

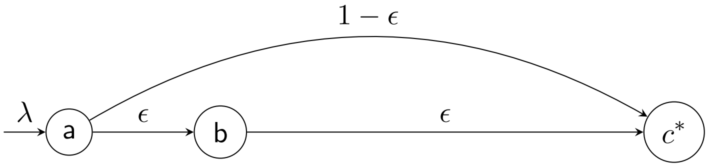
Axiom 3

$$\begin{aligned} p_c &= \sum_b \mathbb{P}(b \mid c) x_{b,c} \\ &= \sum_b \left(\frac{\mu_b P_{bc}^\beta}{\mu_c} \right) \cdot \frac{h_c}{h_b} \cdot (p_b + \beta_b z_b) \end{aligned}$$

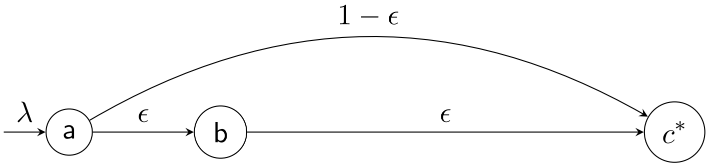
Definition. (Local budget balance)

$$p_c + \beta_c z_c \leq h_c^\beta$$

Revisit previous example



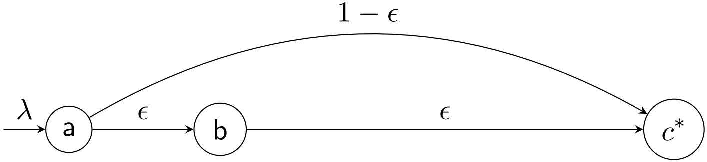
Revisit previous example



Past payments

- $p_a = 0$
- $p_b = z_a \cdot \dfrac{h_b}{h_a} = z_a \cdot \dfrac{\epsilon}{1 - \epsilon + \epsilon^2}$

Revisit previous example



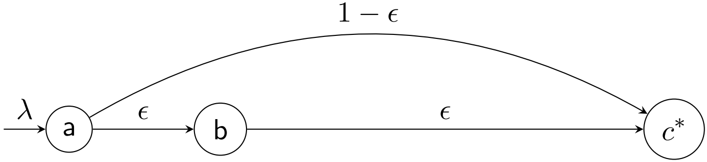
Past payments

- $p_a = 0$
- $p_b = z_a \cdot \dfrac{h_b}{h_a} = z_a \cdot \dfrac{\epsilon}{1-\epsilon+\epsilon^2}$

Budget balance conditions

- $z_a \leq h_a = 1 - \epsilon + \epsilon^2$
- $z_b + \dfrac{\epsilon z_a}{1-\epsilon+\epsilon^2} \leq h_b = \epsilon$

Revisit previous example



Past payments

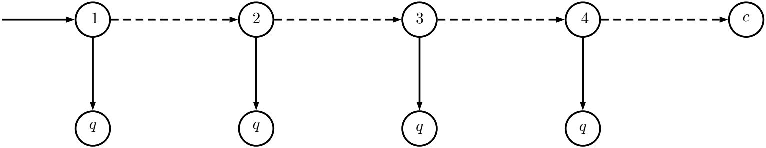
- $p_a = 0$
- $p_b = z_a \cdot \dfrac{h_b}{h_a} = z_a \cdot \dfrac{\epsilon}{1-\epsilon+\epsilon^2}$

Budget balance conditions

- $z_a \leq h_a = 1-\epsilon+\epsilon^2$
- $z_b + \dfrac{\epsilon z_a}{1-\epsilon+\epsilon^2} \leq h_b = \epsilon$

$z_a = 1-2\epsilon+2\epsilon^2-\epsilon^3$ and $z_b = \epsilon^2$ feasible!

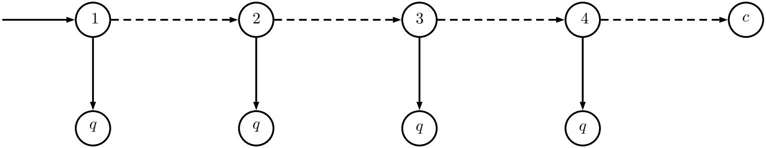
Revisit earlier example



Past payments

$$
\begin{aligned}
p_1 &= 0 \\
p_k &= (p_{k-1} + z_{k-1}) \cdot \frac{h_k}{h_{k-1}} = p_{k-1} + z_{k-1} = \sum_{j<k} z_j
\end{aligned}
$$

Revisit earlier example



Past payments

$$\begin{aligned} p_1 &= 0 \\ p_k &= (p_{k-1} + z_{k-1}) \cdot \frac{h_k}{h_{k-1}} = p_{k-1} + z_{k-1} = \sum_{j<k} z_j \end{aligned}$$

Local budget balance

$$p_k + z_k \leq h_k = 1, \forall k \quad \Leftrightarrow \quad \sum_{k=1}^{n} z_k \leq 1$$

Global budget balance!

Local balance implies global balance!

Theorem. Suppose

$$p_c + \beta_c z_c \ \le \ h_c^\beta, \quad \forall c \qquad \text{(Local budget balance)}$$

Then

$$(\mu^\beta)^\top \mathbf{diag}(\beta) z \ \le \lambda^\top h^\beta. \qquad \text{(Global budget balance)}$$

Proof.

- Use definition of p_c
- Sum all local budget balance constraint
- Careful accounting of slacks

□

Feasible per impression payments z

Channel rationality: $z \geq 0$

Advertiser rationality: global budget balance implied by

- Outcome based cost allocation
- Local budget balance

Fairness

Feasible per impression payments z

Channel rationality: $z \geq 0$

Advertiser rationality: global budget balance implied by

- Outcome based cost allocation
- Local budget balance

Fairness

- IVH with payments: Value V_c at channel c satisfies

$$V = P^{\beta} V + \pi^{*} - \mathbf{diag}(\beta) z$$

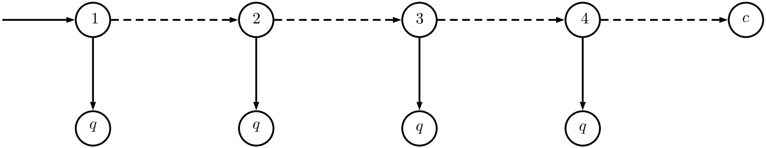
  Reduces to h^{β} when $z \equiv 0$

  $z \leq \Delta V + \pi^{*}$ [No lift $\implies$ No payment]

Does **not** identify z uniquely! Introduce an objective function.

Maximize payments

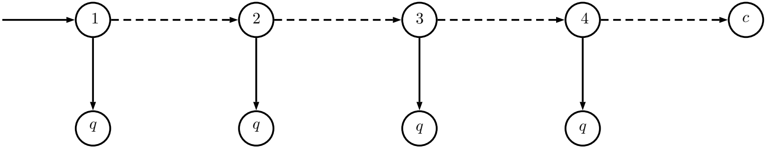
Example



Max payment any payment satisfying global budget constraint $\sum_k z_k \le 1$

Maximize payments

Example



Max payment any payment satisfying global budget constraint $\sum_k z_k \leq 1$

Very fragile!

- $\pi_1^* = \epsilon$: unique optimal solution: $z_1^* = h_1$ and $z_j^* = 0$ for $j > 1$

Nash bargaining solution

Let $\mathcal{Z}$ denote a set of feasible payments for a given network.

$$\mathcal{Z} = \left\{ z \;\middle|\; \begin{array}{l} V = P^{\beta} V + \pi^* - \mathbf{diag}(\beta) z \\ z \leq \Delta V + \pi^* \\ p + \mathbf{diag}(\beta) z \leq h^{\beta} \\ p_c = \sum_b \left(\frac{\mu_b P^{\beta}_{bc}}{\mu_c} \right) \cdot \frac{h_c}{h_b} \cdot (p_b + \beta_b z_b) \end{array} \right\}$$

Nash bargaining solution

Let $\mathcal{Z}$ denote a set of feasible payments for a given network.

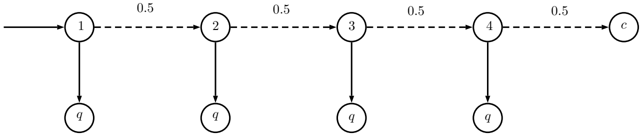
Let f denote a function that maps $\mathcal{Z}$ to an element $z \in \mathcal{Z}$ that satisfies

- Pareto efficiency
- Symmetry
- Invariance to scaling payoffs
- Independence of Irrelevant Alternatives

Then

$$\begin{aligned} f(\mathcal{Z}) = \quad & \max_z \quad \sum_{i=1}^{n} \ln(z_i) \\ & \text{s.t.} \quad z \in \mathcal{Z} \end{aligned}$$

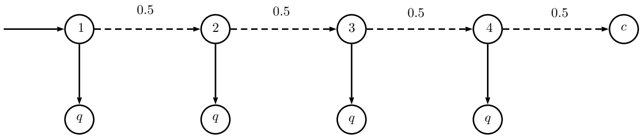
Sensitivity analysis: Path



Nash bargaining solution: $\mu_i \beta_i z_i = \mu_i h_i / n$

	z_1	z_2	z_3	z_4	Total
Max	0.0000	0.0000	0.0000	0.5000	0.0625
Nash	0.0156	0.0313	0.0625	0.1250	0.0625
Nash + IVH	0.0156	0.0312	0.0625	0.1250	0.0625

Sensitivity analysis: Path



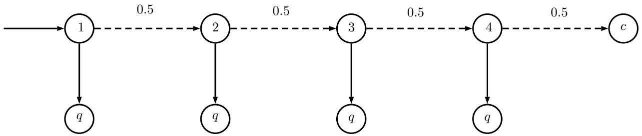
Nash bargaining solution: $\mu_i \beta_i z_i = \mu_i h_i / n$

	z_1	z_2	z_3	z_4	Total
Max	0.0000	0.0000	0.0000	0.5000	0.0625
Nash	0.0156	0.0313	0.0625	0.1250	0.0625
Nash + IVH	0.0156	0.0312	0.0625	0.1250	0.0625

Change $\pi_1^* = 0.1$

	z_1	z_2	z_3	z_4	Total
Max	0.225	0.0	0.0	0.0	0.225
Nash	0.0406	0.0313	0.0625	0.125	0.0875
Nash + IVH	0.0406	0.0313	0.0625	0.125	0.0875

Sensitivity analysis: Path



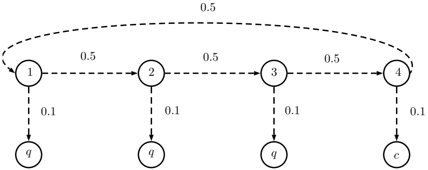
Nash bargaining solution: $\mu_i \beta_i z_i = \mu_i h_i / n$

	z_1	z_2	z_3	z_4	Total
Max	0.0000	0.0000	0.0000	0.5000	0.0625
Nash	0.0156	0.0313	0.0625	0.1250	0.0625
Nash + IVH	0.0156	0.0312	0.0625	0.1250	0.0625

Set $P^{(0)}(3,4) = P^{(1)}(3,4) = 0.5$

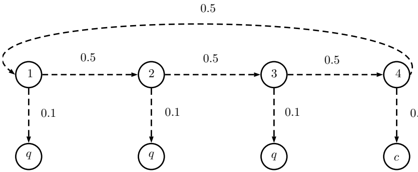
	z_1	z_2	z_3	z_4	Total
Max	0.0000	0.0000	0.0000	0.5000	0.0625
Nash	0.0156	0.0313	0.0625	0.1250	0.0625
Nash + IVH	0.0234	0.0469	0.0000	0.1250	0.0625

Sensitivity analysis: Cycle



	z_1	z_2	z_3	z_4	Total
Max	0.1	0.1	0.1	0.1	0.2
Nash	0.1	0.1	0.1	0.1	0.2
Nash + IVH	0.1	0.1	0.1	0.1	0.2

Sensitivity analysis: Cycle

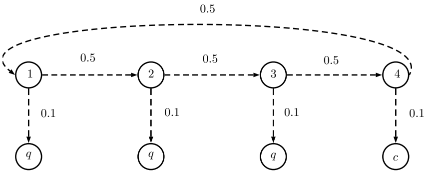


	z_1	z_2	z_3	z_4	Total
Max	0.1	0.1	0.1	0.1	0.2
Nash	0.1	0.1	0.1	0.1	0.2
Nash + IVH	0.1	0.1	0.1	0.1	0.2

Change $\pi_1^* = 0.2$

	z_1	z_2	z_3	z_4	Total
Max	0.1533	0.1067	0.1133	0.1267	0.25
Nash	0.1533	0.1067	0.1133	0.1267	0.25
Nash + IVH	0.1644	0.1000	0.1000	0.1178	0.2411

Sensitivity analysis: Cycle



	z_1	z_2	z_3	z_4	Total
Max	0.1	0.1	0.1	0.1	0.2
Nash	0.1	0.1	0.1	0.1	0.2
Nash + IVH	0.1	0.1	0.1	0.1	0.2

Set $\pi_3^* = 0$ and $P^{(1)}(3,4) = P^{(0)}(3,4) = 0.5$

	z_1	z_2	z_3	z_4	Total
Max	0.0867	0.0733	0.0000	0.1400	0.1500
Nash	0.0867	0.0733	0.0467	0.0933	0.1500
Nash + IVH	0.0860	0.0908	0.0000	0.0527	0.1147

Concluding remarks

Principled framework for attribution

- cost approach
- build on accounting principles
- captures conversion funnel and network effect

Current status

- obtain set of feasible costs per impressions
- avoids limitations of last touch and incremental value heuristics

Further questions

- Alternative criteria for identifying unique z
- Empirical tests

Abhishek, V., Fader, P., and Hosanagar, K. (2012).
Media exposure through the funnel: A model of multi-stage attribution.
Available at SSRN 2158421.

Anderl, E., Becker, I., V Wangenheim, F., and Schumann, J. H. (2014).
Mapping the customer journey: A graph-based framework for online attribution modeling.

Archak, N., Mirrokni, V. S., and Muthukrishnan, S. (2010).
Mining advertiser-specific user behavior using adfactors.
In *Proceedings of the 19th international conference on World wide web*, pages 31–40. ACM.

Berman, R. (2013).
Beyond the last touch: Attribution in online advertising.
Available at SSRN 2384211.

Dalessandro, B., Perlich, C., Stitelman, O., and Provost, F. (2012).
Causally motivated attribution for online advertising.
In *Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy*, page 7. ACM.

Jordan, P., Mahdian, M., Vassilvitskii, S., and Vee, E. (2011).
The multiple attribution problem in pay-per-conversion advertising.
In *Algorithmic Game Theory*, pages 31–43. Springer.

Kireyev, P., Pauwels, K., and Gupta, S. (2013).
Do display ads influence search? attribution and dynamics in online advertising.
Harvard Business School, Boston, MA.

Lewis, R. A. and Rao, J. M. (2014).
The unfavorable economics of measuring the returns to advertising.
Available at SSRN 2367103.

Li, H. and Kannan, P. (2014).
Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment.
Journal of Marketing Research, 51(1):40–56.

Shao, X. and Li, L. (2011).
Data-driven multi-touch attribution models.
In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 258–264. ACM.