Multi-Armed Bandit
Applications to Marketing Experiments

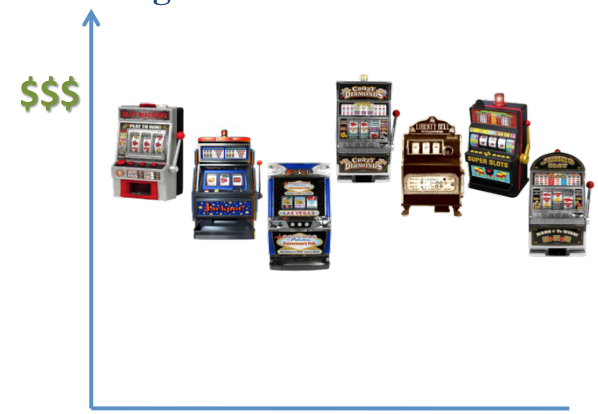
Eric M. Schwartz

ericmsch@umich.edu

ericmichaelschwartz.com

MICHIGAN
ROSS SCHOOL OF BUSINESS

Introducing the multi-armed bandit



MICHIGAN
ROSS SCHOOL OF BUSINESS

2

Introducing the multi-armed bandit

MICHIGAN
ROSS SCHOOL OF BUSINESS

3

Bandit Problem: Key Ideas

Optimization

- Sequentially select actions to maximize reward (minimize regret)

Classic Tradeoff

- Learn vs. Earn (Explore vs. Exploit)
- Long-run vs. Immediate payoff

MICHIGAN
ROSS SCHOOL OF BUSINESS

4

Why is a bandit problem different?

Typical machine learning
- One sample for inference
- "Offline learning"

Bandit problem
- Adaptive sampling
- Active learning
- Partial information / limited feedback
  - Given current data, what data should we collect to optimize an outcome?

Modern business experimentation

Harvard Business Review

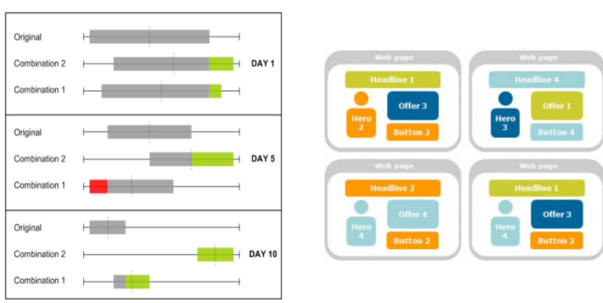
TECHNOLOGY
How to Design Smart Business Experiments
by Thomas H. Davenport
FROM THE FEBRUARY 2009 ISSUE

INNOVATION
A Step-by-Step Guide to Smart Business Experiments
by Eric T. Anderson and Duncan Simester
FROM THE MARCH 2011 ISSUE

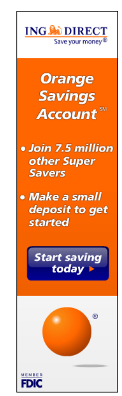
INNOVATION
The Discipline of Business Experimentation
by Stefan Thomke and Jim Manzi
FROM THE DECEMBER 2014 ISSUE

APT

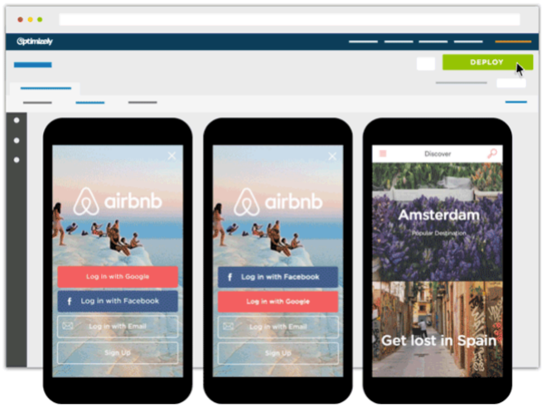
A/B testing



Current examples: Online ads

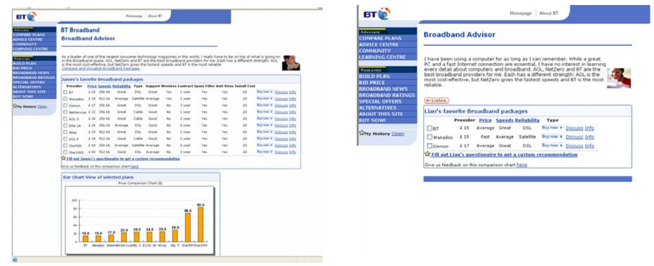


Current examples: App design



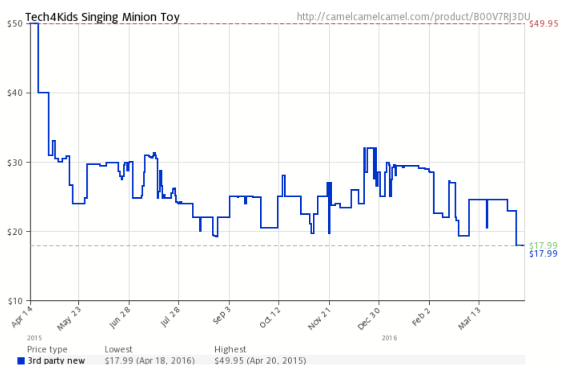
Current examples: Website style

Comparison of Two Morphs for a Website Advisor



(a) General content, large-load, graphical morph (b) Focused, small-load, verbal morph

Current examples: Pricing



Current examples: Medial clinical trials

THE UNIVERSITY OF TEXAS
MD Anderson
~~Cancer~~ Center

- Patients have life threatening disease
  - Patients receive one of two different treatments in a clinical trial
- Ethical dilemma: Collective v Every Individual
  - Learn the best treatment (Clinical Research)
  - Treat the patient better (Clinical Practice)

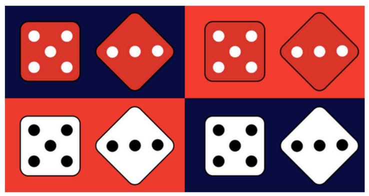
Current examples

Netflix Shares a Decade of A/B Test Learning

2015 SXSW Interactive Session Spotlight: A/B Testing at Scale – Minimizing UI Complexity

Big experiments: Big data's friend for making decisions

By Eytan Bakshy on Thursday, April 3, 2014 at 1:31pm



MICHIGAN
ROSS SCHOOL OF BUSINESS

Current examples

Optimizely monetate®

"We're in a world now where you can't wait a whole day to get your data. This has been a huge boon for the growth of our business."

ONE KINGS LANE

"Testing is a way of life for us. We're increasing our testing resources, and building testing into our long-term, strategic plans."

AVALANCHE TECHNOLOGY GROUP

MICHIGAN
ROSS SCHOOL OF BUSINESS

Big Data?

The Atlantic

The Big Data Boom Is the Innovation Story of Our Time

AN EXPERIMENT EVERY SECO

ERIK BRYNJOLFSSON AND ANDREW MCAFEE | NOV 21 2011, 9:50 AM ET

Science has been dominated by the experimental approach for nearly 400 years. Running controlled experiments is the gold standard for sorting out cause and effect. But experimentation has been difficult for businesses throughout history because of cost, speed and convenience. It is only recently that businesses have learned to run real-time experiments on their customers. The key enabler was the Web.

Hal Varian: Google "is running on the order of 100-200 experiments on any given day, as they test new products and services, new algorithms and alternative designs."

Greg Linden (Amazon): "Constant, ubiquitous experimentation is the most important thing."

of 100-200 experiments on any given day, as they test new products and services, new algorithms and alternative designs. An iterative review process aggregates findings and frequently leads to further rounds of more targeted experimentation.

MICHIGAN
ROSS SCHOOL OF BUSINESS

Real-time experiment methods in practice

- A/B testing (A/B/C/…/n)
- Multivariate testing (MVT)
- Stopping rules
- Multi-armed bandit experiments
  - Adaptive allocation of observations

MICHIGAN
ROSS SCHOOL OF BUSINESS

Example

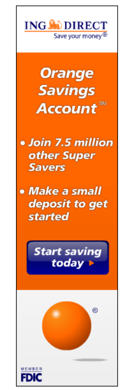
Customer Acquisition via Display Advertising Using Multi-Armed Bandit Experiments

Schwartz, Bradlow, and Fader (2016)

Marketing Science, forthcoming

MICHIGAN
ROSS SCHOOL OF BUSINESS

Which ad will bring in the most new customers?

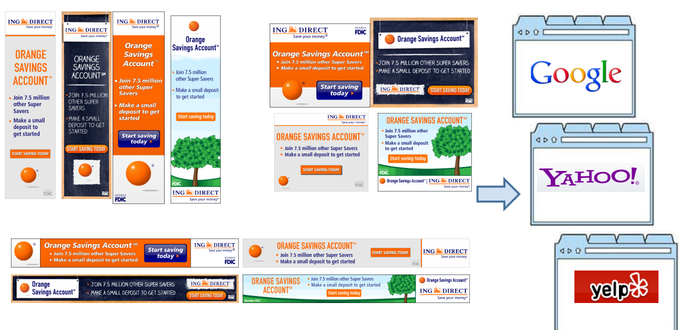


Always be testing and learning

Always be earning while learning

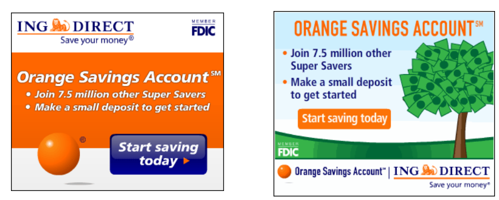
MICHIGAN
ROSS SCHOOL OF BUSINESS

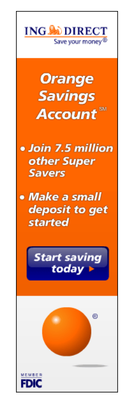
How should we allocate impressions across many ads (served on many websites) to acquire more customers?



MICHIGAN
ROSS SCHOOL OF BUSINESS

Always be testing and learning





Always be earning while learning

MICHIGAN
ROSS SCHOOL OF BUSINESS

20

Typical analytics for A/B tests or MVT

Combinations (24) | Page Sections (2) | Download: XML CSV TSV | Print

Disable | All Combinations (24) | Key: Winner Inconclusive Loser

Combination	Status	Est. conv. rate	Chance to Beat Orig.	Observed Improvement	Conv./Visitors
Original	Enabled	8.26% ± 0.5%	—	—	1088 / 13167
Top high-confidence winners. Run a follow-up experiment »					
Combination 11	Enabled	11.6% ± 0.6%	100%	40.6%	1504 / 12947
Combination 7	Enabled	10.3% ± 0.6%	100%	24.0%	1340 / 13073
Combination 3	Enabled	9.80% ± 0.6%	99.7%	18.7%	1277 / 13025
Combination 10	Enabled	9.23% ± 0.6%	95.9%	11.7%	1203 / 13031
Combination 8	Enabled	9.03% ± 0.6%	91.6%	9.28%	1178 / 13046
Combination 9	Enabled	8.77% ± 0.6%	81.8%	6.10%	1111 / 12672
Combination 6	Enabled	8.64% ± 0.5%	75.3%	4.58%	1108 / 12822

MICHIGAN
ROSS SCHOOL OF BUSINESS

Google Content Experiments

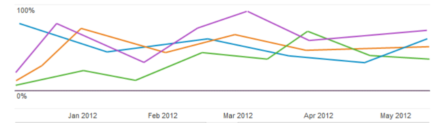
% of total experiment visits : 100.00%

Explorer

Conversions | Site Usage | Goal Set 1 | Goal Set 2 | Goal Set 3 | Goal Set 4 | Ecommerce

Experiment running - no winner yet

Conversion Rate vs. Select a metric | Day | Week | Month



331 visits
161 days of data
100% visitors included

Status:
No winner yet - Experiment still running

Primary Dimension: Variation

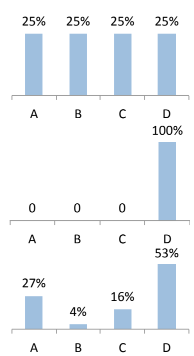
Plot Rows

Variation	Experiment Visits	Conversions	Conversion Rate	Compare to Original	Probability of Outperforming Original
Original	64	34	52.31%	0%	0.00%
Variation 2	70	39	55.71%	7%	66.11%
Variation 3	96	45	46.88%	-10%	25.14%
Variation 1	101	3	2.97%	-94%	0.00%

MICHIGAN
ROSS SCHOOL OF BUSINESS

Managing the multi-armed bandit

1. Static, balanced design
   (equal allocation)
2. Adaptive, "greedy" methods
   (winner take all)
3. Adaptive, randomized
   (smooth allocation)



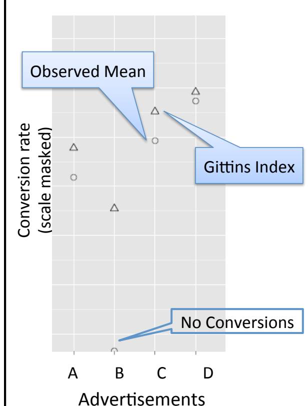
Agarwal et al. 2008; Auer et al. 2002; Bertsimas and Mersereau 2007; Lai 1987; Scott 2010; Rusmevichientong and Tsitsiklis 2010

Gonul and Shi 1998; Gonul and ter Hosftede 2006; Hauser et al. 2009; Montoya et al. 2011; Simester et al. 2006; Sun et al. 2006; Urban et al. 2014

MICHIGAN
ROSS SCHOOL OF BUSINESS

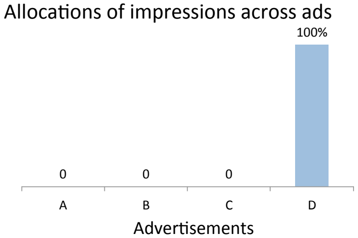
23

Managing the multi-armed bandit



Winner-take-all policies
Observed mean ("Greedy")
Gittins Index
Upper Confidence Bound (UCB)

Allocations of impressions across ads



MICHIGAN
ROSS SCHOOL OF BUSINESS

24

List of bandit algorithms

- Learn-then-earn
- epsilon-Greedy
- Logit/Boltzman transform (Softmax)
- Exponential Weight (EXP3)

- *Gittins index*
- *Upper Confidence Bound (UCB)*
- *Thompson Sampling*

MICHIGAN
ROSS SCHOOL OF BUSINESS

25

The problem we can solve with Gittins index

Stylized bandit problem

Independent actions.
Not attribute-based.

x_k are K-dimensional indicator vectors

$\theta = (\mu_1, \ldots, \mu_K)$

$J = 1$

No hierarchical structure.
One context.

$M_{it} = 1$ for all t

No batched decisions.
One-at-a-time.

$\mu_k = \mathrm{E}_f(y|x_k)$

$Y_{kt} \sim f(y|x_k)$

$Y_{kt} = R(A_t, \mu)$

Objective function

$$\max_{A_t \in 1,\ldots,K} \int_{\mu_1} \cdots \int_{\mu_K} \mathrm{E}_f \left\{ \sum_{t=1}^{\infty} \gamma^t R(A_t, \mu) \right\} p(\mu_1) \ldots dp(\mu_K) d\mu_1 \ldots d\mu_K$$

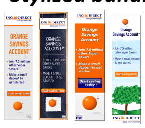
Solved! Exactly optimal policy = Gittins index policy

Gittins 1979, 1989; Gittins and Jones 1974; Tsitsitklis 1986

26

The problem we can solve with Gittins index

Stylized bandit problem *(ex. Bernoulli bandit with beta priors)*



$$a_{kt} = a_{k0} + \sum_{\tau=1}^{t} y_{k\tau} \text{ and } b_{kt} = b_{k0} + \sum_{\tau=1}^{t} (m_{k\tau} - y_{k\tau})$$

$$\Pr(Y_{kt} = 1 | a_{kt}, b_{kt}) = E_{p(\mu_k)}(\mu | a_{kt}, b_{kt}) = \frac{a_{kt}}{a_{kt} + b_{kt}}$$

$$V(a_{kt}, b_{kt}, \gamma) = \max \left\{ \frac{G_{kt}}{1-\gamma}, \right.$$

K separable "one-and-a-half-armed bandit"

$$[1 + \gamma V(a_{kt}+1, b_{kt}, \gamma)] \frac{a_{kt}}{a_{kt}+b_{kt}}$$
$$\left. + [0 + \gamma V(a_{kt}, b_{kt}+1, \gamma)] \frac{b_{kt}}{a_{kt}+b_{kt}} \right\}$$

•G is the Gittins index, which is exactly optimal value.
•Represents "certainty equivalent," "option value," etc.
•Policy: At any state, play the action with the highest Gittins index.

27

Gittins index and UCB

Stylized bandit problem

Gittins index (exactly optimal solution)

$$\max \left\{ \frac{G_{kt}}{1-\gamma}, \underbrace{\frac{a_{kt}}{a_{kt}+b_{kt}}}_{\text{Expected Immediate Reward}} + \gamma \underbrace{\left[V(a_{kt}+1, b_{kt}, \gamma) \frac{a_{kt}}{a_{kt}+b_{kt}} + V(a_{kt}, b_{kt}+1, \gamma) \frac{b_{kt}}{a_{kt}+b_{kt}} \right]}_{\text{Expected Option Value for Learning}} \right\}.$$

Upper confidence bound (asymptotically optimal)

$$\mathrm{UCB1} = \underbrace{\frac{1}{n_k(t)} \sum_{i=1}^{n_k(t)} y_{ki}}_{} + \underbrace{\sqrt{\frac{2\log(t)}{n_k(t)}}}_{}$$

$$\mathrm{Regret}_t = \sum_{\tau=1}^{t} (\mu_* - \mu_{A_t})$$

Agarwal et al. 2008; Auer et al. 2002; Brezzi and Lai 2002; Lai 1987

28

The problems we can solve

Attribute-based bandit problem

(still no unobserved context heterogeneity, no batching)

29

The problems we can solve

Attribute-based bandit problem

(still no unobserved context heterogeneity J=1, no batching)

Use spillover learning to leverage similarity of actions (regression framework) and predict performance of not yet chosen actions (a la conjoint).

$$\text{UCB-GLM}_{kt} = \text{link}^{-1}(x_k'\beta_t) + \|x_k\|_{Q_{t-1}^{-1}}\rho(t)$$

$$\|\beta - \hat{\beta}_t\|_{Q_t} \leq \rho^*(t)$$

$$\|x_k\|_{Q_t^{-1}} = \sqrt{x_k' Q_t^{-1} x_k}$$

Linear regression structure

$$\text{“} (X'X)^{-1} \text{”}$$

$$Q_t = \sum_{\tau=1}^{t-1} x_{A_\tau} x_{A_\tau}'$$

Nests the UCB1 algorithm and any myopic GLM (regression model) without learning.

Minimizes “regret” with high probability (Dani et al. 2008; Filippi et al. 2010; Rusmevietong and Tsitsiklis 2010)

30

Life is more complicated…

- Natural extensions and complications in testing
  - Large set of candidates to compare (e.g., creative content, display ads)
  - Very rare events (e.g., transactions or customer acquisitions via display ads)
  - Batches of decisions (e.g., “chunky” allocations)
  - Different contexts (e.g., websites, customer segments differ)
  - Long-run customer value (e.g., impact of actions beyond one transaction)

MICHIGAN
ROSS SCHOOL OF BUSINESS

The problem we want to solve

How should we **allocate** impressions across many ads served on many websites to acquire more customers?

Hierarchical attribute-based bandit

MICHIGAN
ROSS SCHOOL OF BUSINESS

32

Overview of solution

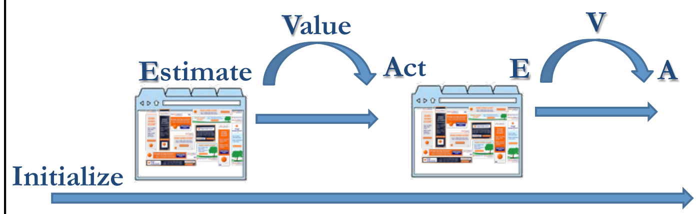
- Improve efficiency of online advertising
- Broaden the class of earn-and-learn problems through managing multi-armed bandit with
  – heterogeneity (hierarchical structure)
  – attribute structure (non-independent actions)
  – batched decisions
  – rare events

  by extending the use of Thompson Sampling.
- Document performance of methods
  – Field experiment implementing policy
  – Counterfactual simulations

MICHIGAN
ROSS SCHOOL OF BUSINESS

Adaptive experiment timeline

Data are conversions and impressions for each website-ad combo.

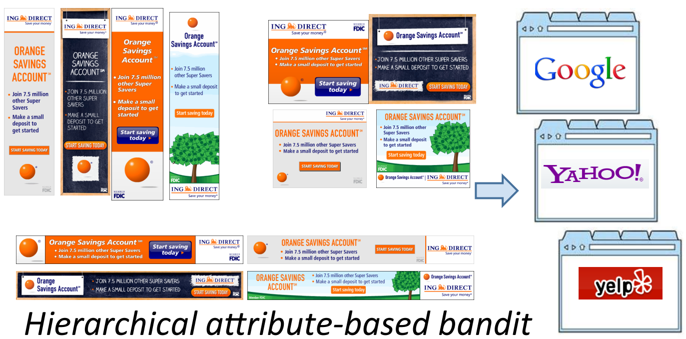


Time of the experiment

MICHIGAN
ROSS SCHOOL OF BUSINESS

The problem we want to solve

How should we **allocate** impressions across many ads served on many websites to acquire more customers?

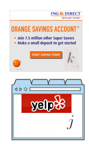


Hierarchical attribute-based bandit

MICHIGAN
ROSS SCHOOL OF BUSINESS

Preliminaries for optimization problem

Hierarchical attribute-based *K-armed bandit with J contexts and* ***batching***



$$\{y_{jk1}, \ldots, y_{jkt}\} \sim f(y|x_k, \beta_j) \forall j, k$$

$$x_k = (x_{k1}, \ldots, x_{kd})$$

Attribute-based

Hierarchical structure

$$\beta_j \sim g(\bar{\beta}, \Sigma) \forall j \qquad \theta = (\{\beta_j\}_1^J, \bar{\beta}, \Sigma)$$

$$\mu_{jk} = \mathrm{E}_f(y|x_k, \beta_j) = \text{link}^{-1}(x_k' \beta_j) \forall j, k$$

Batched decisions

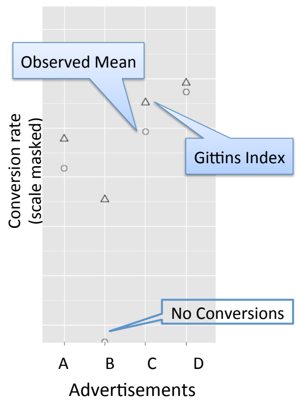
$$1 = \sum_{k=1}^{K} w_{jkt} \forall j, t \text{ batch of } M_{jt} \text{ observations}$$

$$R(w_{jt}, \beta_j) = \textstyle\sum_{k=1}^{K} Y_{jkt}$$

$$\mathrm{E}_f R(w_{jt}, \beta_j) = \sum_{k=1}^{K} w_{jkt} M_{jt} \mu_{jk} = \sum_{k=1}^{K} w_{jkt} M_{jt} \text{link}^{-1}(x_k' \beta_j) \forall j.$$

$$\text{Regret}_T(w) = \sum_{t=1}^{T} \sum_{j=1}^{J} M_{jt} \left(\mu_{j*} - \sum_{k}^{K} w_{jkt} \mu_{jk} \right)$$

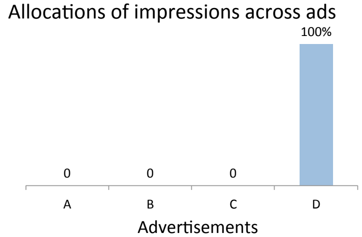
Managing the multi-armed bandit



Winner-take-all policies

- Observed mean ("Greedy")
- Gittins Index
- Upper Confidence Bound (UCB)

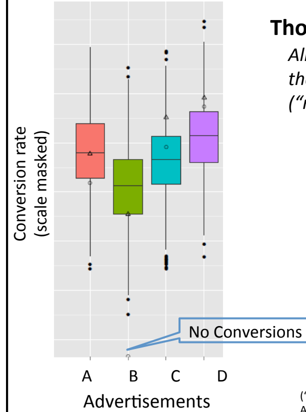
Allocations of impressions across ads



MICHIGAN ROSS SCHOOL OF BUSINESS

37

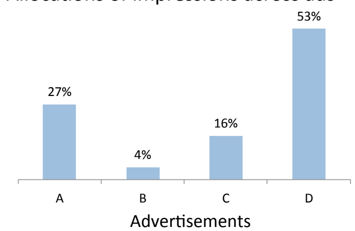
Managing the multi-armed bandit



Thompson Sampling

Allocate resources to the action in proportion to the probability that is the best action. ("randomized probability matching")

Allocations of impressions across ads



("Asymptotic optimality," "minimax regret," "Bayes risk," "minimum entropy") Agrawal and Goytal 2012; Berry 2004; Chapelle and Li 2012; Granmo 2010; Kaufmann et al. 2012; May et al 2011; Ortega and Braun 2013; Russo and Van Roy 2014; Scott 2010; Thompson 1933.

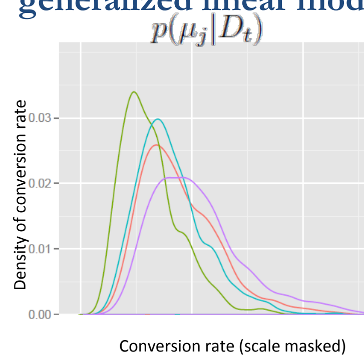
MICHIGAN ROSS SCHOOL OF BUSINESS

38

Managing the multi-armed bandit

Online Simulator for Bayesian Bandits with Thompson Sampling
https://googledrive.com/host/0B2GQktu-wcTiWDB2R2t2a2tMUG8/

Thompson Sampling (TS) with hierarchical generalized linear model (HGLM)



$[\mu_{1j}, \ldots, \mu_{Kj}]$

Heterogeneity is novel to TS approach for bandit problems

$$\mu_{j*} = \max\{\mu_{j1}, \ldots, \mu_{jK}\}$$

$$w_{jk} = \Pr\left(\mu_{jk} = \mu_{j*}|D_t\right)$$

MICHIGAN ROSS SCHOOL OF BUSINESS

Agrawal and Goyal 2012; Berry 2004; Chapelle and Li 2012; Granmo 2010; May et al 2011; Russo and Van Roy 2014; Scott 2010; Thompson 1933

40

Thompson Sampling (TS) with hierarchical generalized linear model (HGLM)

$$\mu_{j*} = \max\{\mu_{j1}, \ldots, \mu_{jK}\}$$

website-specific "winner" among ads

$$w_{jk} = \Pr\left(\mu_{jk} = \mu_{j*} | D_t\right)$$

website-specific probabilities of each ad being the winner

$$w_{jk} = \int_{\mu_j} \mathbf{1}\left\{\mu_{jk} = \mu_{j*} | \mu_j\right\} p(\mu_j | D_t) d\mu_j$$

average over uncertainty in each ad's conversion rate

$$w_{jk} \approx \hat{w}_{jk} = \frac{1}{G} \sum_{g=1}^{G} \mathbf{1}\left\{\mu_{jk}^{(g)} = \mu_{j*}^{(g)} | \mu_j^{(g)}\right\}$$

$$(m_{1,j,t+1}, \ldots, m_{K,j,t+1}) = (\hat{w}_{j1t}, \ldots, \hat{w}_{jKt}) M_{j,t+1}$$

"match" proportional allocations to probabilities

MICHIGAN
ROSS SCHOOL OF BUSINESS

41

Field experiment: summary and scope

- **700** million impressions
- **80** media placements (websites)
- **15** publishers
- Conversion rates were within industry standards (between **1 and 10 out of 1 million** impressions)
- For each website, **12** ads are described by attributes (**4** ad concepts **x 3** ad sizes)

MICHIGAN
ROSS SCHOOL OF BUSINESS

42

Field experiment: ad attribute structure

- For each website:

  4 ad concepts and **3** ad sizes

Conversion rate indexed as percent of average

	Ad A	Ad B	Ad C	Ad D	
Tall 160x600	117	88	99	176	
Square 300x250	107	72	151	114	
Wide 728x90	115	92	66	72	
All Sizes	**112**	**80**	**100**	**105**	100

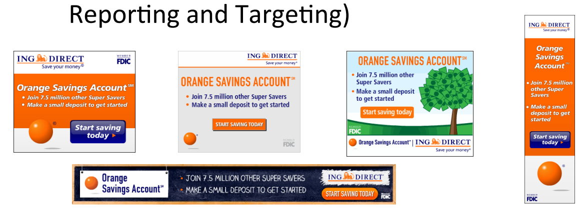
Best concept can differ by size

MICHIGAN
ROSS SCHOOL OF BUSINESS

43

How was the field experiment implemented?

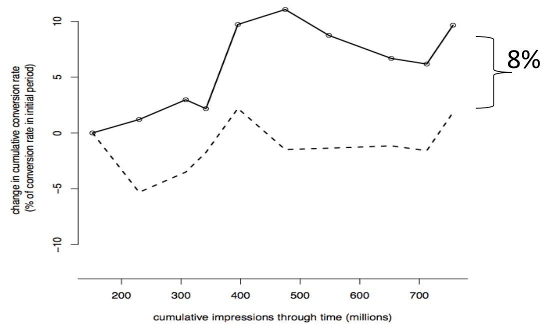
- Timing
  - Update every 6 days for 61 days in 2012
- Allocation probabilities are "rotation weights"
  - Receive data and upload weights directly to Google DoubleClick DART (Dynamic Advertising Reporting and Targeting)



MICHIGAN
ROSS SCHOOL OF BUSINESS

44

Live field experiment: TS *versus* Balanced



- Test: Adaptive experiment (**TS**)
- Control: Static **balanced** experiment

ING DIRECT

MICHIGAN
ROSS SCHOOL OF BUSINESS

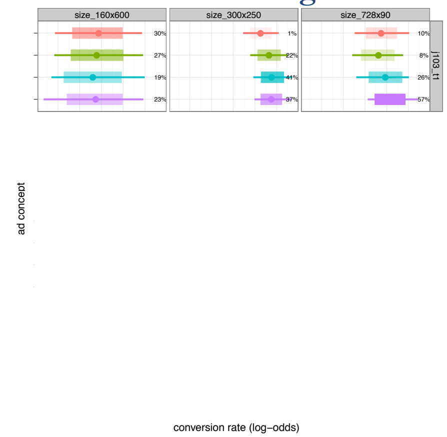
45

How does TS learn for heterogeneous allocation?

*4 concepts within 3 sizes, 2 "websites"

*Boxplots show posterior of conversion rates. (right is better)

*Darker coloring shows more allocation of impressions.

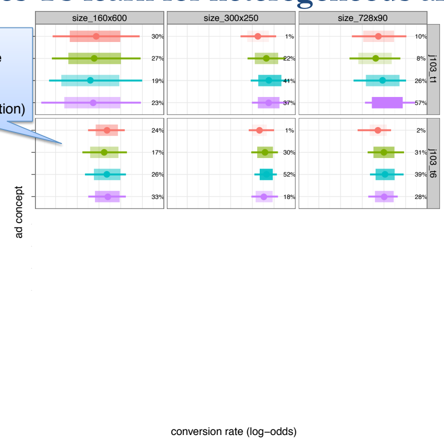


"website" j

MICHIGAN
ROSS SCHOOL OF BUSINESS

How does TS learn for heterogeneous allocation?

Learning each ad conversion rate over time for a website (uncertainty reduction)

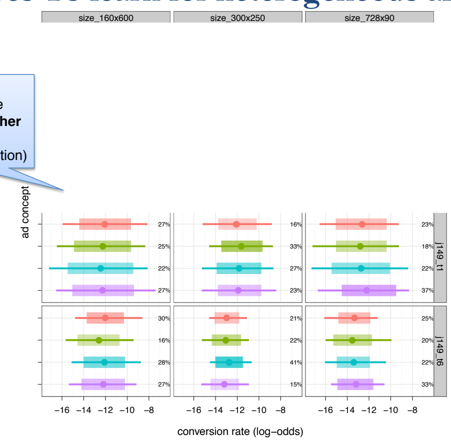


"website" j

MICHIGAN
ROSS SCHOOL OF BUSINESS

How does TS learn for heterogeneous allocation?

Learning each ad conversion rate over time for **another** website (uncertainty reduction)

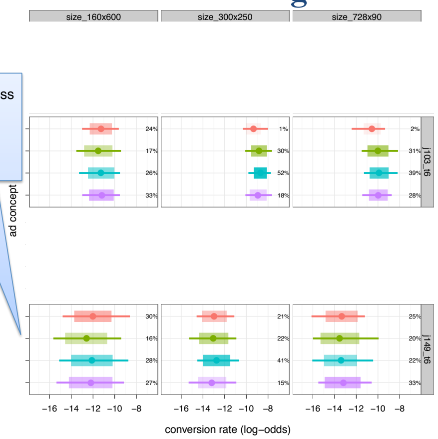


"website" j'

MICHIGAN
ROSS SCHOOL OF BUSINESS

How does TS learn for heterogeneous allocation?

Heterogeneity across websites: the same ads have different conversion rate and different allocations



Counter-factual simulations based on field experiment

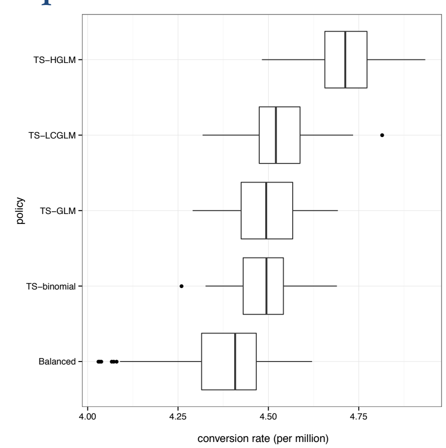
Counter-factual simulations based on field experiment

Run any policy to sequentially reallocate impressions given budget and timing constraints.

1. Compare TS-based policies with different models
   -attribute-based: logit (GLM), latent class (LCGLM)
2. Compare greedy and epsilon-greedy policies
   -Play the winner, always exploit (greedy)
   -Randomize: Explore (ε%) and exploit (1- ε%) (epsilon-greedy)
3. Compare intuitive test-rollout policies
   -Balanced experiment, then play the winner (explore, then exploit)

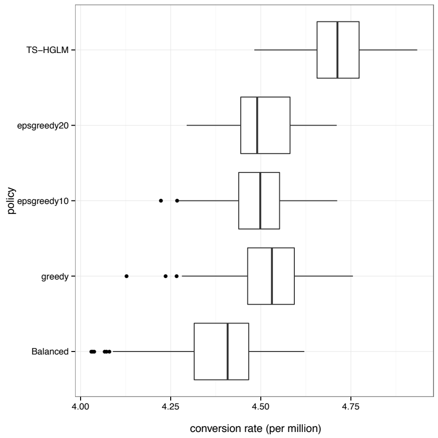
51

TS-based policies with different models



52

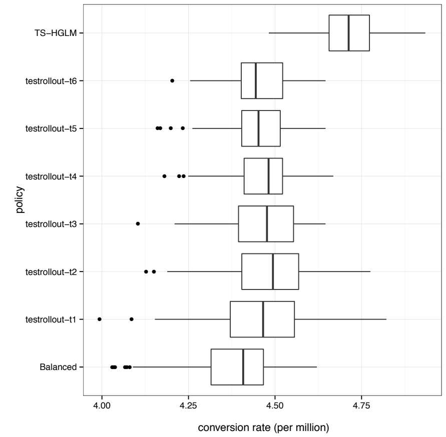
Greedy and epsilon-greedy policies



MICHIGAN
ROSS SCHOOL OF BUSINESS

53

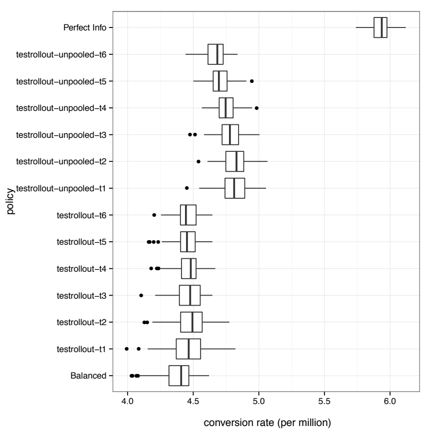
Test-rollout policies



MICHIGAN
ROSS SCHOOL OF BUSINESS

54

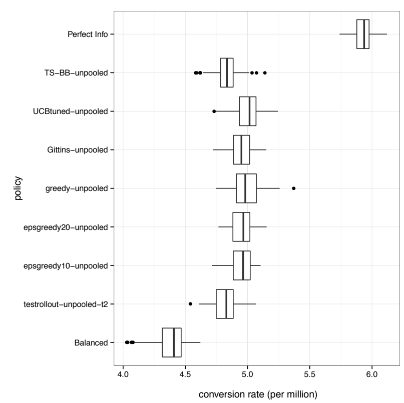
Model heterogeneity vs. Allocation rule



MICHIGAN
ROSS SCHOOL OF BUSINESS

55

Heterogeneity matters most unpooled and partially pooled beat pooled



MICHIGAN
ROSS SCHOOL OF BUSINESS

56

Summary of counterfactuals

- Accounting for heterogeneity across context improves performance
- Hierarchical model (partial pooling) with bandit outperforms standard (pooled) models with bandit
- Even more flexible forms of heterogeneity can beat hierarchical bandit
- The modeling story matters more than particular bandit allocation rule...
  ... so get the story right!

MICHIGAN
ROSS SCHOOL OF BUSINESS

57

Summary

- Large scale real-time adaptive field experiment and simulations show benefits of Thompson Sampling with appropriate model.
- Firms that experiment adaptively and systematically for continuous improvement should be "earning while learning."

MICHIGAN
ROSS SCHOOL OF BUSINESS

58

Summary

- Large scale real-time adaptive field experiment and simulations show benefits of Thompson Sampling with appropriate model.
- The proposed hierarchical bandit beats standard models used with bandit algorithms.
- Firms that experiment adaptively and systematically for continuous improvement should be "earning while learning."

MICHIGAN
ROSS SCHOOL OF BUSINESS

59

Future directions

- Dynamic (robust) firm pricing over many periods can improve when combining bandit methods with econ theory
- Market research survey techniques that adapt to respondents answers can more efficiently utilize its sample with bandit learning
- Consider lifetime value when exploring and exploiting new sources of customer acquisition

MICHIGAN
ROSS SCHOOL OF BUSINESS

60

Takeaways

- Active learning and adaptive sampling
- Learn vs. Earn tradeoff
- Many algorithms (Gittins, UCB, Thompson, egreedy) from in many disciplines
- Real-world problems motivate complicated versions with “bells and whistles”

- Always be earning while learning

MICHIGAN
ROSS SCHOOL OF BUSINESS

61

THANK YOU!

ericmsch@umich.edu
ericmichaelschwartz.com

MICHIGAN
ROSS SCHOOL OF BUSINESS

62

Additional Details on Multi-Armed Bandit Optimization Methods

MICHIGAN
ROSS SCHOOL OF BUSINESS

64

Dynamic Robust Pricing with Multi-Armed Bandits

- How should we set price with limited information about demand to maximize profit over time?

MICHIGAN
ROSS SCHOOL OF BUSINESS

65

Dynamic Robust Pricing Overview

Research objective: provide a scalable method for price experimentation with computer science algorithms and economic theory

- Pricing with incomplete information
- Learning demand and price experimentation
- Prices are bandit arms
- Dynamic pricing
- Robust dynamic pricing
- Combing machine learning and pricing

MICHIGAN
ROSS SCHOOL OF BUSINESS

66

Pricing with incomplete information

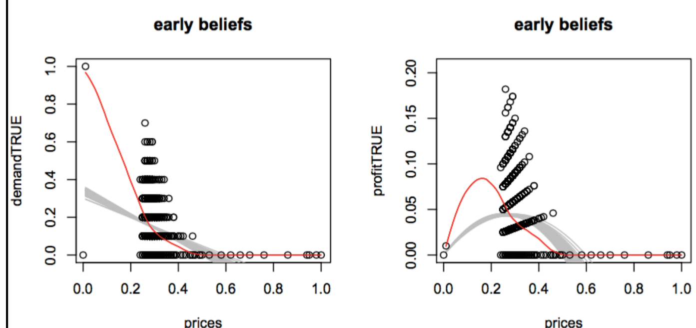
- Consider a firm objective of pricing a new product
  - Firm must set a price
  - Will assume consumers have stable preferences
- Limited data available

- **How can a manager set a price?**
  - Experiment!

MICHIGAN
ROSS SCHOOL OF BUSINESS

67

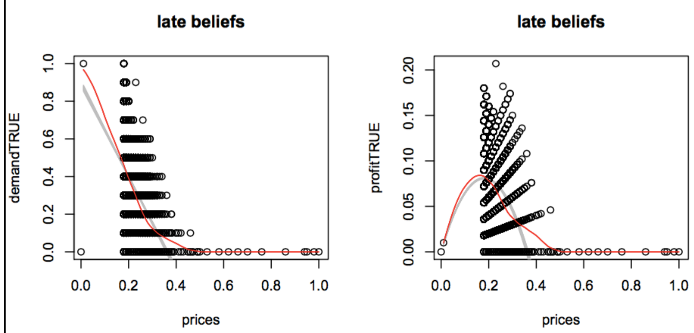
Learning: assuming linear demand



MICHIGAN
ROSS SCHOOL OF BUSINESS

68

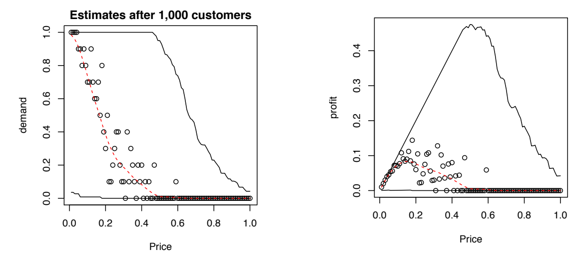
Learning: assuming linear demand



MICHIGAN
ROSS SCHOOL OF BUSINESS

69

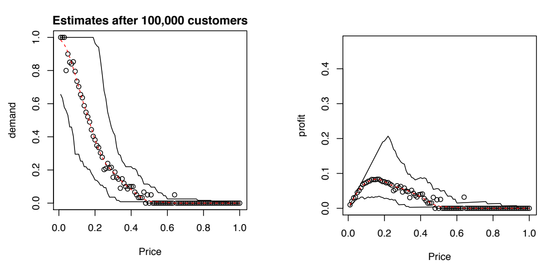
Learning: no parametric assumption



MICHIGAN
ROSS SCHOOL OF BUSINESS

70

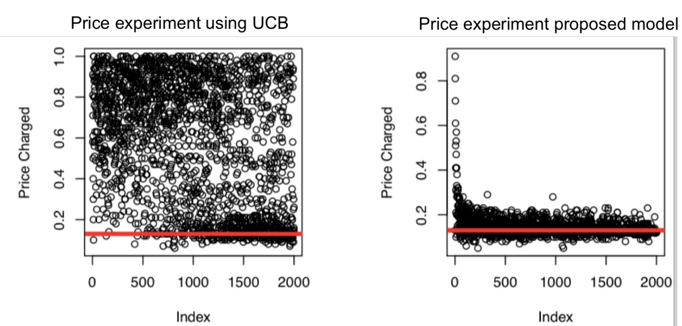
Learning: no parametric assumption



MICHIGAN
ROSS SCHOOL OF BUSINESS

71

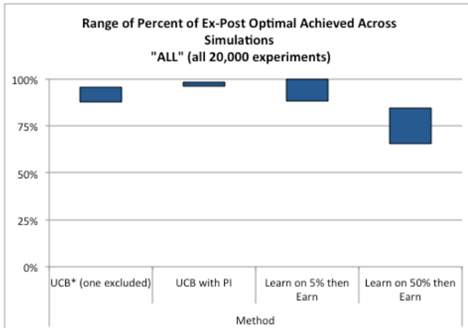
Converging to best price faster



MICHIGAN
ROSS SCHOOL OF BUSINESS

72

Proposed bandit pricing best controls risk



MICHIGAN
ROSS SCHOOL OF BUSINESS

73

Future directions

- Dynamic (robust) firm pricing over many periods can improve when combining bandit methods with econ theory
- Market research survey techniques that adapt to respondents answers can more efficiently utilize its sample with bandit learning
- Consider lifetime value when exploring and exploiting new sources of customer acquisition

MICHIGAN
ROSS SCHOOL OF BUSINESS

74

How does TS 'optimally' solve the problem?

For independent actions, no heterogeneity, one-at-a-time decisions …

- Gittins index is the optimal certainty equivalent of an uncertain arm (1979)

But… that's not our problem!

- Explore/Exploit is balanced by sampling from full distribution of beliefs
- TS is asymptotically optimal in maximizing cumulative reward (i.e., minimax regret shrinks in log time).

Agrawal and Goyal 2012; Berry 2004; Chapelle and Li 2012; Granmo 2010; Kaufmann et al. 2012; May et al 2011; Russo and Van Roy 2014; Scott 2010; Thompson 1933

MICHIGAN
ROSS SCHOOL OF BUSINESS

75

Thompson Sampling (TS) with hierarchical generalized linear model (HGLM)

$\mu_{j*} = \max\{\mu_{j1}, \ldots, \mu_{jK}\}$ website-specific "winner" among ads

$w_{jk} = \Pr(\mu_{jk} = \mu_{j*} | D_t)$ website-specific probabilities of each ad being the winner

$$w_{jk} = \int_{\mu_j} \mathbf{1}\{\mu_{jk} = \mu_{j*} | \mu_j\} p(\mu_j | D_t) d\mu_j$$

average over uncertainty in each ad's conversion rate

$$w_{jk} \approx \hat{w}_{jk} = \frac{1}{G} \sum_{g=1}^{G} \mathbf{1}\left\{\mu_{jk}^{(g)} = \mu_{j*}^{(g)} | \mu_j^{(g)}\right\}$$

$$(m_{1,j,t+1}, \ldots, m_{K,j,t+1}) = (\hat{w}_{j1t}, \ldots, \hat{w}_{jKt}) M_{j,t+1}$$

"match" proportional allocations to probabilities

MICHIGAN
ROSS SCHOOL OF BUSINESS

76