Online Advertising Incrementality Testing And Experimentation: Industry Practical Lessons

Joel Barajas, Narayan Bhamidipati
Yahoo Research, Verizon Media
joel.barajas@verizonmedia.com
narayanb@verizonmedia.com

James G. Shanahan
Church and Duncan Group Inc., UC Berkeley
James.Shanahan@gmail.com

Abstract

Online advertising has historically been approached as user targeting and ad-to-user matching problems within sophisticated optimization algorithms. As the research area and ad tech industry have progressed over the last couple of decades, advertisers have increasingly emphasized the causal effect estimation of their ads (aka incrementality) using controlled experiments (or A/B testing). Even though observational approaches have been derived in marketing science since the 80s including media mix models, the availability of online advertising personalization has enabled the deployment of more rigorous randomized controlled experiments with millions of individuals.

These evolutions in marketing science, online advertising, and the ad tech industry have posed incredible challenges for engineers, data scientists, and marketers alike. With low effect percentage differences (or lift) and often sparse conversion rates, the development of incrementality testing platforms at scale suggests tremendous engineering challenges in the measurement precision and detailed implementation. Similarly, the correct interpretation of results addressing a business goal within the marketing science domain requires significant data science and experimentation research expertise. All these challenges on the ongoing evolution of the online advertising industry and the heterogeneity of its sources (social, paid search, native, programmatic, etc).

In the current tutorial, we propose a practical, grounded view in the incrementality testing landscape, including:

- The business need
- Solutions in the literature
- Design and choices in the development of incrementality testing platform
- The testing cycle, case studies, and recommendations to effective results delivery
- Incrementality testing evolution in the industry

We will provide first-hand lessons on developing and operationalizing such a platform in a major combined DSP and ad network; these are based on running tens of experiments for up to two months each over the last couple of years.

Intended Audience

Practitioners and Researchers in the field of online advertising and marketing.
General knowledge of online advertising, familiarity with a/b testing is a plus.

Previous offering(s)

Econometric Analysis and Digital Marketing: How to Measure the Effectiveness (link) from the Web Conference 2013.
This tutorial was given at the time ad effectiveness testing was gaining popularity in the industry from an econometric perspective. Our current tutorial shows the eight-year evolution of the methods, real-world practices, and operationalization of those practices.

Learning outcomes

Topics Reviewed:
1. Experiment design and A/B testing
2. Causal inference
3. Ad tech architectures
4. Business use cases in advertising and marketing analytics
5. Practical recommendations to successful significant incrementality experiments

Participants will:
1. Identify and formulate key approaches to measuring the effectiveness of online advertising.
2. Execute relevant statistics for hypothesis testing, power analysis in experiment planning and simulate experiment scenarios.
3. Be able to define key ingredients of an operational incrementality testing platform and their trade-offs.
4. Understand the business need for incrementality testing.
5. Identify the necessary conditions to increase the likelihood of successful test given minimum detectable lift, conversion type, test duration among others.

Take-away material
1. **Cheatsheet of statistical tools**, including python notebooks and spreadsheets for statistical power analysis
2. **Summary of recommendations handout** document for operationalization addressing the challenges and trade-offs of operationalizing testing
3. **Categorization of business use cases handout** document addressing typical needs and focus of incrementality testing

Bios:

Joel Barajas, Sr Research Scientist, Yahoo Research (Verizon Media)

Joel has over 11 years of experience in the online advertising industry with research contributions at the intersection of Ad tech, Marketing Science, and Experimentation. He has experience with Ad load personalization and experimentation in a publisher marketplace. Within Marketing Data Science orgs, he has supported regular budget allocation and Media Mix Models in multi-channel advertising. With a PhD dissertation focussed on ad incrementality testing, his published work has appeared in top outlets including INFORMS Marketing Science Journal, ACM CIKM, ACM WWW, SIAM SDM. He led the science development and marketing analytics of the incrementality testing platform in a multidisciplinary team. He currently oversees most incrementality tests in Verizon Media ad network (previously yahoo!) and DSP (previously AOL advertising.com). Joel also leads the science development in CTV and linear TV measurement modeling. He holds a B.S. (with honors) in Electrical and Electronics Engineering from the Tecnológico de Monterrey, and a PhD in Electrical Engineering (with emphasis on statistics) from UC Santa Cruz.

Narayan Bhamidipati, Sr Director of Research, Yahoo Research (Verizon Media)
Narayan has over 14 years of experience in Computational Advertising and Machine Learning. He currently leads a team of researchers focused on providing state-of-the-art ad targeting solutions to help ads be more effective and relevant. This includes creating various contextual targeting products to reduce the company's reliance on user profiles and help improve monetization in a more privacy aware world. Alongside that, Narayan ensures that the user profile based ad targeting products continue to improve despite the decline of tracking data. In addition, Narayan is keen on developing the most accurate ad effectiveness measurement platform which would help the company attract more revenue by proving the true value of the ad spend on our platforms. He holds B.Stat(Hons), M.Stat and PhD(CS) degrees, all from the Indian Statistical Institute, Kolkata.

Dr. James G. Shanahan, Church and Duncan Group Inc. and UC Berkeley.
Dr. James G. Shanahan has spent the past 30 years developing and researching cutting-edge artificial intelligence systems, splitting his time between industry and academia. For the academic year 2019-2020, Jimi held the position of Rowe Professor of Data Science at Bryant University, Rhode Island. He has (co) founded several companies that leverage AI/machine learning/deep learning/computer vision in verticals such as digital advertising, web search, local search, and smart cameras. Previously he has held appointments at AT&T (Executive Director of Research), NativeX (SVP of data science), Xerox Research (staff research scientist), and Mitsubishi. He is on the board of Anvia, and he also advises several high-tech startups including Aylien, ChartBoost, DigitalBank, LucidWorks, and others. Dr. Shanahan received his PhD in engineering mathematics and computer vision from the University of Bristol, U. K.
Jimi has been involved with KDD since 2004 as an author, as a tutorial presenter, and as a workshop co-chair; he has actively been involved as a PC/SPC member over the years also.

Tutorial Outline

Part 1: The basics: context and challenges

- The problem
  - Online Advertising spend trends between performance and brand
  - Big picture problem: quarterly/yearly budget allocation
  - Budget allocation practices based on financial models
  - The need for testing combined with industry attribution practices
- How channel-level testing fits within other forms of testing
  - Real-time decision making in targeting engines
  - Tactic testing: A/B testing with last-touch attribution
  - Multi-cell testing A/B testing + Incrementality testing
  - CMO decision making at the end of the quarter/semester/year
    - Marketing Channel Managers: social, paid search, display, radio, tv, etc
    - Cross-channel Managers: overall channel performance and recommendations for budget allocations
- Incrementality Testing: Business Use cases
  - Advertiser joining new partners
    - “Test the waters first, then join with full budgets”
  - Testing to calibrate and rebase financial models
    - Last-touch attribution multipliers
    - Media Mix Models among different channels
  - The Marketing component
    - Growth marketing: acquisition, signups, quality signups, LTV
    - CRM marketing: sales, churn, retention, product launch
    - Testing over different signals in the conversion funnel

Part 2: Incrementality Testing: concepts, solutions and literature

- Incrementality Testing: Literature and Industry practices
  - Placebo Based Testing: practice and issues
    - Mechanics
    - User-level randomization
    - Subject to targeting bias: not blind to serving engine
  - Intention to treat Testing
    - Mechanics
    - Practices to identify treatment groups
    - The challenge with reachable but never exposed groups
    - Pitfalls
  - Ghost Ads testing idea
    - Concepts
    - Practical Execution
    - Ghost Bidding as Intention to treat experiment
    - Ghost Ads within ad networks

        - Benefits: experiment precision
- Estimation Frameworks
    - Econometric causality
        - Structural Equation Model
    - Potential Outcomes Causal Framework
        - Units, treatments and potencial outcomes
    - Pitfalls
        - Endogenous variables and bias
        - Post-treatment biases
        - Sympson's paradox in experiment analysis

Part 3: From concept to production: platform building, challenges, case studies

Building the experiment platform

- The Building Journey
    - Engage stakeholders
    - The business case
    - Simulation tools
    - Getting everybody onboard: 80% technical 20%
- The identity graph and treatment groups
    - Cookie-based experiments
        - One of the earlier approaches in web
        - Standard in display advertising within DSPs
        - Problems: cookie deletion, one per device, etc
        - Recommendations from literature
    - Device-based experiments
        - Standard in companies measuring cross-channel ads in open exchanges
        - Reliable within ad networks for in-app ads
        - Problems: id reset, normalization, hashing, etc
        - Recommendations from literature
    - Logged-in users based experiments
        - Standard approaches ids are email addresses
        - Common within the wallet gardens
        - Cross-device experiment execution and attribution
        - As reliable as product experimentation
    - Household-level experiments
        - Identity graphs providers based on identity IP clustering
        - Problems: IP churn and reset, traveling, moving, etc
        - Recommendations
- User holdout design within modern Ad tech serving systems
    - The hashing functions
        - Prior user groups randomization and sticky assignments
        - Live user hashing
    - The challenge with targeting and scoring algorithms
        - ML-based targeting as a source of bias

        - Re-targeting and prospective strategies as sources of bias
        - Interactions among campaign lines
    - How to avoid targeting bias
        - The PSA campaign design under regular ML targeting optimization
        - Double-blind experiment designs
    - The role of look-back windows in last-touch attribution engines
        - Last-touch attribution relevance during experiments
        - Any-touch attribution to resemble product experiment
        - Look-back window constraints and meaning
        - Recommendations to assess different look-back windows
- Data Logging and Analysis
    - Data logging
        - The experiments units and the identity graph power
        - Defining the experiment eligibility event
        - Defining the experiment response and interaction with standard attribution
    - Analysis
        - Potential Outcomes Causal Framework Principles
        - Experiment units as a source of variability among individuals
        - Hypothesis testing and confidence intervals

Part 4: Deployment at Scale: test cycle and case studies

Experiment execution cycle

- Experiment Design and Planning
    - Recap on the solution ingredients
        - Ads versus no ads in testing
        - Controlled experiment design
        - User groups and sizes
        - Incremental ad effects given all other advertising forms
    - Statistical power analysis
        - Hypothesis testing
        - Forecast
        - Conversion rates and user universe
    - Budget estimations and recommendations
    - Experiment duration
    - Expectation management based on the conversion response
- Intervention Execution
    - Control/treatment identification and logging
    - Sanity checks and A/A tests
- Experiment Tracking and Metrics
    - Typical performance marketing performance metrics
    - Experiment monitoring
    - Effect trends and holidays
- End of experiment readout
    - The key metrics

    - Key recommendations or insights
        - Insights
        - Recommendations
    - User response under seasonality
    - Comparisons with last-touch attributed value

Case Studies

- Insurance quotes and comparison with post-click conversions
    - Quotes as a conversion metric
    - The effect of irrelevant holidays
- Online food ordering revenue: CRM versus New audiences
    - Revenue as conversion metric
    - The fundamental difference between CRM and new users
- Online acquisition signup
    - Acquisition as a metric
    - Excluding existing users
    - The effect of relevant holidays

Part 5: Emerging trends: identity challenges, industry trends and solutions

Advertisers Testing without Ad Networks

- Spend as experiment intervention
    - Cost curves and channel elasticities
- Methodologies: Time series based testing
    - The concept and contrast with observational studies
    - Difference-in-Difference based approaches
    - Synthetic controlled based approaches

Geo-testing

- Geo units specification
    - Challenges with DMAs
    - Challenges with zip codes
- Geo unit treatment assignment
    - Correlation analysis
    - A/A test as experimental design tool

Challenges with user ids

- The cookiless world
    - User group randomization as proxy of user levels
    - Short-term first-party data as proxy for advertiser’s conversions response
    - Challenges and emerging problems

References

The need for Incrementality Testing Solutions

1. Gordon, Brett R., et al. "A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook." *Marketing Science* 38.2 (2019): 193-225.

2. Kireyev, Pavel, Koen Pauwels, and Sunil Gupta. "Do display ads influence search? Attribution and dynamics in online advertising." *International Journal of Research in Marketing* 33.3 (2016): 475-490.
3. Li, Hongshuang, and P. K. Kannan. "Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment." *Journal of Marketing Research* 51.1 (2014): 40-56.
4. Chan, David, et al. "Evaluating online ad campaigns in a pipeline: causal models at scale." *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2010

Incrementality Testing Solutions

5. Barajas, Joel and Narayan Bhamidipati. "Incrementality Testing in Programmatic Advertising: Enhanced Precision with Double-Blind Designs" *Proceedings of the 30th International Web Conference.* 2021. To appear
6. Johnson, Garrett A., Randall A. Lewis, and Elmar I. Nubbemeyer. "Ghost ads: Improving the economics of measuring online ad effectiveness." *Journal of Marketing Research* 54.6 (2017): 867-884.
7. Barajas, Joel, et al. "Experimental designs and estimation for online display advertising attribution in marketplaces." *Marketing Science* 35.3 (2016): 465-483.
8. Lewis, Randall A., Justin M. Rao, and David H. Reiley. "Here, there, and everywhere: correlated online behaviors can lead to overestimates of the effects of advertising." *Proceedings of the 20th international conference on World wide web*. 2011.

Causal Inference

9. Rubin, Donald B. "Causal inference using potential outcomes: Design, modeling, decisions." *Journal of the American Statistical Association* 100.469 (2005): 322-331.
10. Frangakis, Constantine E., and Donald B. Rubin. "Principal stratification in causal inference." *Biometrics* 58.1 (2002): 21-29.
11. Imbens, Guido W., and Donald B. Rubin. "Bayesian inference for causal effects in randomized experiments with noncompliance." *The annals of statistics* (1997): 305-327.
12. The Importance of Being Causal, Iavor Bojinov, Albert Chen, and Min Liu, HDSR 2.3, (2020)

Operationalization and Practical Recommendations

13. Barajas, Joel and Narayan Bhamidipati. "Incrementality Testing in Programmatic Advertising: Enhanced Precision with Double-Blind Designs" *Proceedings of the 30th International Web Conference.* 2021. To appear
14. Kohavi, Ron, Diane Tang, and Ya Xu. *Trustworthy online controlled experiments: A practical guide to a/b testing*. Cambridge University Press, 2020.
15. Lewis, Randall A., and Justin M. Rao. "The unfavorable economics of measuring the returns to advertising." *The Quarterly Journal of Economics* 130.4 (2015): 1941-1973.

Geo-testing and Synthetic Control and Identity Challenges

16. Barajas, Joel, Tom Zidar, and Mert Bay. "Advertising Incrementality Measurement using Controlled Geo-Experiments: The Universal App Campaign Case Study." AdKdd (2020).
17. Blake, Thomas, Chris Nosko, and Steven Tadelis. "Consumer heterogeneity and paid search effectiveness: A large-scale field experiment." *Econometrica* 83.1 (2015): 155-174.
18. Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American statistical Association* 105.490 (2010): 493-505.
19. Barajas, Joel, et al. "Dynamic effects of ad impressions on commercial actions in display advertising." *Proceedings of the 21st ACM international conference on Information and knowledge management*. 2012.
20. Lin, Tesary, and Sanjog Misra. "The identity fragmentation bias." *arXiv preprint arXiv:2008.12849* (2020).

Strategies to encourage audience engagement

1. Ice breakers with A/B tests the audience might have done in the past
2. Interactive life demonstration of statistical power analysis
3. Ask questions to connect with test challenges the audience have faced
4. Real case studies