Privacy Aware Experiments without Cookies

Shiv Shankar
University of Massachusetts, USA

Ritwik Sinha
Adobe Research, USA

Saayan Mitra
Adobe Research, USA

Moumita Sinha
Adobe Inc. , USA

Viswanathan Swaminathan
Adobe Research, USA

Sridhar Mahadevan
Adobe Research, USA

ABSTRACT

Consider two brands that want to jointly test alternate web experiences for their customers with an A/B test. Such collaborative tests are today enabled using *third-party cookies*, where each brand has information on the identity of visitors to another website. With the imminent elimination of third-party cookies, such A/B tests will become untenable. We propose a two-stage experimental design, where the two brands only need to agree on high-level aggregate parameters of the experiment to test the alternate experiences. Our design respects the privacy of customers. We propose an estimater of the Average Treatment Effect (ATE), show that it is unbiased and theoretically compute its variance. Our demonstration describes how a marketer for a brand can design such an experiment and analyze the results. On real and simulated data, we show that the approach provides valid estimate of the ATE with low variance and is robust to the proportion of visitors overlapping across the brands.

CCS CONCEPTS

• **Applied computing** → **Marketing**; *Electronic commerce*; • **Security and privacy** → **Privacy protections**;

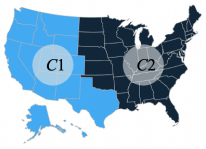
KEYWORDS

Advertising effects, Cookie-less internet, treatment effect

ACM Reference Format:
Shiv Shankar, Ritwik Sinha, Saayan Mitra, Moumita Sinha, Viswanathan Swaminathan, and Sridhar Mahadevan. 2018. Privacy Aware Experiments without Cookies. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (WSDM '23)*. ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

A/B testing (randomized experiment) is the gold standard for optimizing customer experiences on the web. Cookies have historically played an important role in ensuring that the same customer receives a consistent (or *sticky*) experience, thus ensuring the validity of the A/B test. Consider the A/B testing scenarios, where different brands are collectively testing alternate customer experiences. Some examples include, brands under a hotel chain, different franchises under a sporting league, related clothing brands jointly owned by a holding company, or multiple departments of a government. Thus far, such A/B tests have relied on the *third-party cookie*, which requires brand 1's cookie being available on brand 2's website. This is the only mechanism to ensure that the same individual receives the same experience across both web properties.



		Ctrl: Buy	Trt: Buy
Brand	Cluster	Proportion	Proportion
1	$C1$	α	$1-\alpha$
1	$C2$	$1-\alpha$	α
2	$C1$	$1-\alpha$	α
2	$C2$	α	$1-\alpha$

$$\hat{\text{ATE}} = \frac{1}{2\alpha - 1}(\alpha(\bar{Y}_1^{C1} - \bar{Y}_2^{C2}) + (1-\alpha)(\bar{Y}_2^{C1} - \bar{Y}_1^{C2}))$$

Figure 1: Proposed experimental design for two brands to estimate the average treatment effect (ATE) without third party cookies. The brands only agree on a definition of clusters ($C1$ & $C2$), the two treatments (orange and blue "buy" buttons), and agree to randomize in proportions (α & $(1-\alpha)$) swapped across brands and clusters.

A recent trend on the web has been the increased focus customer privacy. This is embodied in laws like the General Data Protection Regulation (GDPR)[1], which require an explicit permission from the visitor for tracking web sessions. A second manifestation of this trend is that all major browser ecosystems have taken steps to discontinue third-party cookies [**?**]. This presents significant challenges for brands that want to jointly optimize customer experiences [18].

Formally, this is our question of interest. Consider a binary treatment deployed across two websites, for example, Brand 1 and Brand 2 have jointly decided to optimize the experiences on their websites (e.g., the discount offer, color of the buy button, layout of the website, and so on). Let's call the treatments 1 and 2. A user may visit one or both websites. In the ideal case a user who was assigned to the group which experiences treatment 1 on the first website, will also see treatment 1 on the second website. However, in the case where the identity of the user cannot be confirmed across the two websites (due to a lack of third-party cookies), the user is not guaranteed to see the same treatment across both websites. This leads to four potential treatment exposures, namely Y_{11}, Y_{21}, Y_{12}, and Y_{22} (first and second positions denoting the treatments at the first and second websites). The true treatments of potential

WSDM '23, Feb 27– March 03, 2023, Singapore

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

[1] https://gdpr-info.eu/

outcomes of interest are Y_{11} and Y_{22}. The question of interest is whether the treatment has any effect; which in the world with cookies corresponds to the effect $E[Y_{11} - Y_{22}]$. This problem is easily solved if the websites could share information about which users were exposed to which treatment; while retaining independent decision/allocation process.

However, it is not clear whether this effect can be estimated without sharing of such individual user level information. In our work we present a multi-stage randomization design which allows the estimation of the desired effect without sharing any user-level information. Figure 1 describes our proposed design. The two brands only agree on two aggregate parameters of the test: 1) the two treatments being tested, (2) a notion of clusters (divide the population into $C1$ and $C2$, e.g., geographies, time, device types), and (3) the nuisance parameter α ($\neq 1/2$). No individual or identity data needs to be shared between the brands. First, we show that our proposed average treatment effect (ATE) is unbiased and compute its variance theoretically. Next, we further propose a way of performing regression adjustment, which further helps statistical power by using attributes of the web visitors. Finally, on experiments in simulated and real world web experiment data, we show that our estimate has lower bias than the naive ATE (difference of treatment means). In our demo, we show how a marketer for a brand can use our design a two-stage test and analyze the results.

2 RELATED WORK

The problems of using cookie-level identifiers on the web as as a proxy for the individual's true identity is well studied [2, 4, 8, 15, 22]. But looking beyond cookies as the identity in digital marketing is less well studied [26]. The Privacy Sandbox [3] is an initiative with proposals tackling privacy related challenges for ad-targeting, delivery and measurement. These include proposals like FLoC and TOPICS, but they are all early proposals subject to changes and uncertainty.

Resolving the user's identity can overcome most measurement and attribution [24, 25] issues, and considerable research has gone into stitching fragmented user behaviour [6, 10, 12, 23]. But these strategies rely on using generic features such as IP to represent the same human user, or use more detailed but private data. The first persists privacy concerns and leads to inaccurate stitching, while the second relies on first party data from walled gardens and precludes any cross-ecosystem analysis. Industrial consortiums like "The TradeDesk Unified ID 2.0" and "Advertising ID Consortium", are building their own unique identifiers based on people-based identifiers (Email ID, Device). Some firms such as Zillow also create custom segments from their first-party data and allow advertisers access. However concerns about scalability and possible future regulation, means this is not sustainable long term strategy. As such Kamena [11] propose for greater use of media mix modeling as a complementary approach to user-level attribution.

Interference related problems have been well studied in the literature [1, 9]. But these assume strong restrictions on the structure of spillover. Recently some work has focused on how to account for general interference [17, 19, 27] . However all of these methods rely on complete knowledge of interference structure which is not possible with our setup.

Another problem related to disappearance of cookies is that of identity fragmentation [4, 16]. Lin and Misra [14] demonstrate treatment effect attenuation in presence of identity fragmentation. Coey and Bailey [5] also provide a debiasing estimator for cookie-level estimates under similar assumptions. But both of these primarily target he loss of first-party cookies.

Our approach is instead based on conducting parallel experimentation followed by stratified aggregation. Instead of constructing user-level records, this approach works by running multiple experimentation at stratified cohorts and constructing user group-level records (aggregation). These are then shared across channels/brands, to compute the desired estimate. Our proposed estimation method is unbiased, but with a higher variance. However by relaxing the need for constructing user-level data, it retains privacy, works without further assumptions and can achieve a wider coverage. Our work borrows the idea of using multi-cluster experimentation from the work of Hudgens and Halloran [9]. Hudgens and Halloran [9] which proposed a two-stage randomization procedure to account for interference. Our scenario however is different in that the same unit is receives multiple treatments instead of having exposure to treatments of other units. Furthermore, we do not make exposure level assumptions.

3 PROBLEM STATEMENT

Let's say that our two websites have users in set $\mathbb{A}$ and $\mathbb{B}$. Note that $\mathbb{A} \cap \mathbb{B} \neq \emptyset$; so some users will visit both. When third party cookies are available, these set of users will get a consistent view (in terms of experience/offers/ads etc); but without cookies this cannot be guaranteed, and a user may be exposed to multiple treatments.

For each user $u_i \in \mathbb{A} \cup \mathbb{B}$ we have an associated feature vector x_i. Each user $u_i \in \mathbb{A}$ can be exposed to treatments 1 or 2 on website 1 . Similarly, each user $u_i \in \mathbb{B}$ can be exposed to treatments 1 or 2 on website 2. If a user $u_i \in \mathbb{A} \cap \mathbb{B}$ then they have exposure to treatments from both websites. On the other hand those in $\mathbb{A} \Delta \mathbb{B}$ are only exposed to one website and we label their other exposure on the other website as "0". Since the two brands collaborate without sharing data, our approach takes the perspective of one of the two brands (since they can only analyze visitors on their website), without loss of generality, let's consider the first brand. For users in set $\mathbb{A}$ we have 6 potential outcome variables, i.e., $Y_{1,0}, Y_{2,0}, Y_{1,1}, Y_{1,2}, Y_{2,1}, Y_{2,2}$. Of these variables only one is observed for each user based on the treatment they received and whether they visited both websites. The problem is estimating the average treatment effect of shifting from treatment 2 to 1. [2]

The challenges in estimation arises because of two main issues a) each website can only control allocation of treatments to their users; and b) the identity of the shared users is unknown. Whenever a user u_i visits, the websites can only access the features x_i; and choose a treatment without information on whether $u_i \in \mathbb{A} \cap \mathbb{B}$ or how the other website might allocate treatment. Equivalently phrased we know for a user u_i of website 1 and allocated to treatment 1 that we are observing one of $Y_{1,.}$, $Y_{1,1}$ or $Y_{1,2}$ but not which of these. On the other hand in standard causal inference we know for each observation unit, whether the observed outcome is Y_1 or Y_2.

[2] We make two standard assumptions in treatment effect literature, strong ignorability and positivity. A greater discussion about these is present in the supplementary material

Let us make this more explicit: For a group which has been allocated treatment 1 by website 1, the expected average outcome is given by:

$$\mathbb{E}[\tilde{Y}_1] = (1-p)\mathbb{E}[Y_{1,0}] + p(\alpha\mathbb{E}[Y_{1,1}] + (1-\alpha)\mathbb{E}[Y_{1,2}]).$$

Here, $\tilde{Y}_1$ is the random variable denoting the observed outcome of a visitor to website 1 who is randomized to treatment 1, p is the fraction of users who are shared and hence visit both websites, while α is the fraction of these shared users who receive treatment 1 on the second website as well. Every user has the probability $1-p$ of only visiting website 1 and hence consistently receives treatment 1. For these users the average outcome is $\mathbb{E}[Y_{1,0}]$. For the rest of the users (who are p fraction of the population), an α fraction of them are allocated to treatment 1 by website 2 (and hence are exposed to treatment pair 1, 1). The observed outcome on these users is $\mathbb{E}[Y_{1,1}]$. Similarly a $1-\alpha$ fraction of the shared users receive treatment 2 from website and produce the average outcome $\mathbb{E}[Y_{1,2}]$. The observed average effect is the probability weighted combination of all the contributions.

Furthermore by symmetry between the treatments, we can write a similar equation for the average outcome of group allocated treatment 2 by website 1. The observed advantage of treatment 1 over treatment 2, i.e., $\mathbb{E}[\tilde{Y}_1] - \mathbb{E}[\tilde{Y}_2]$ (which is also the standard treatment effect estimate) is given by:

$$(1-p)\mathbb{E}[Y_{1,0} - Y_{2,0}] + p(\alpha\mathbb{E}[Y_{1,1} - Y_{2,1}] + (1-\alpha)\mathbb{E}[Y_{1,2} - Y_{2,2}])$$

Next, we analyze the treatment effect if one could track the users and provide them a consistent experience. Every user has the probability $1-p$ of only being on website 1 and hence receiving consistently either treatment 1 or 2. For these users the treatment effect is $\mathbb{E}[Y_{1,0} - Y_{2,0}]$. The rest of the users; who are p fraction of the population; consistently receive treatment pair $(1,1)$ or $(2,2)$. The corresponding effect is $\mathbb{E}[Y_{1,1} - Y_{2,2}]$. The average treatment effect is the population weighted combination of the two contributions.

$$\text{TE} = (1-p)\mathbb{E}[Y_{1,0} - Y_{2,0}] + p(\mathbb{E}[Y_{1,2} - Y_{2,2}])$$

It is clear from the desired treatment effect and the observed treatment effect are mismatched due to contributions from the cross treatment outcomes $Y_{2,1}$ and $Y_{1,2}$. Moreover it is clear that the mismatch increases with p the fraction of shared users.

In the next section we present our method of estimating the true treatment effect.

4 METHOD

We assume that we are able to make more than 1 macroclusters with a suitable mix of treatment allocation. This is similar to the multi-stage randomization technique of Hudgens and Halloran [9]. Each user first gets assigned to a macrocluster which corresponds to one of the treatment allocation strategies. Then within each such macrocluster randomization is done for allotment of treatment. One possible way to solve this is to do stratified randomization with consistence across platforms/websites. More specifically for each user u_i we put them in a category/macrocluster based on their features x_i, and this function or mapping the users to macroclusters is shared across all websites. This can be achieved via something similar to FLoC or Topics API which allows advertisers partial view of the users preferences. Note that privacy is still maintained here; since no website shares user-specific information with the other website.

Recall that the goal of this exercise is to estimate the following estimand of interest: $(1-p)\mathbb{E}[Y_{1,0} - Y_{2,0}] + p(\mathbb{E}[Y_{1,3} - Y_{2,4}])$. This reflects the the following notation. The treatments options at website 1 are T_1 and T_2, and those at website 2 are T_3 and T_4. The notation $E[Y_{1,0}]$ denotes the expectation of the outcome in visitors who have seen T_1 at first website but have not visited website 2. If we are in the most likely scenario where $T_1 = T_3$ and $T_2 = T_4$, our estimand boils down to $(1-p)\mathbb{E}[Y_{1,0} - Y_{2,0}] + p(\mathbb{E}[Y_{1,1} - Y_{2,2}])$. In the interest of keeping our solution generic, we are sticking to notation with four treatments T_1, T_2, T_3, and T_4.

Using this macro-level aggregation of user information, one can do randomization with the following approach. Lets assume that we have only 2 macroclusters C_1 and C_2. In the first cluster the allocation ratio of the treatments T3 and T4 as chosen by the second website is $\alpha \neq 0.5$; while in the second cluster the allocation ratio is $1-\alpha$. We depict the observed average outcomes for cluster Ci and treatment j by the variable $\bar{Y}_j^{Ci}$ (an estimate of $\mathbb{E}(\tilde{Y}_j^{Ci})$). Then given our setting we have the four observed outcomes $\bar{Y}_1^{C1}, \bar{Y}_2^{C1}, \bar{Y}_1^{C2}, \bar{Y}_2^{C2}$.

Our estimator of the treatment effect is given by:

$$\hat{\text{TE}} = \frac{1}{2\alpha - 1}(\alpha\bar{Y}_1^{C1} + (1-\alpha)\bar{Y}_2^{C1} - (1-\alpha)\bar{Y}_1^{C2} - \alpha\bar{Y}_2^{C2}) \quad (1)$$

Claim 1. *$\hat{TE}$ is an unbiased estimate of the treatment effect TE*

A proof of this claim along with and estimate of the variance of this estimator can be found in Appendix A.

4.1 Covariate Adjustment

Assuming the randomization is perfect, the earlier treatment effect estimate is unbiased unconditionally. However when there is an imbalance in respect of some covariate between the groups, adjusting for the baseline effect leads to a more efficient estimator. Next we provide a regression based method to improve the efficiency of our earlier estimator.

Algorithm 1 Covariate Adjustment Algorithm

Input: Vector of outcomes $Y_1^{C1}, Y_2^{C1}, Y_1^{C2}, Y_2^{C1}$ allocation α
Output: Covariate adjusted treatment effect $\hat{\text{ATE}}_{\text{cov}}$

(1) Let $Z_i^{Cj} = \delta Y_i^{Cj}$ i.e. it is 1 if Y corresponds to treatment 1and 0 otherwise
(2) Fit OLS($[Y_1^{C1}, Y_2^{C2}] \sim X + 1 + [Z_1^{C1}, Z_2^{C2}]$)
(3) Let β_1 be coefficient of Z as estimated in previous step
(4) Fit OLS($[Y_2^{C1}, Y_1^{C2}] \sim X + 1 + [Z_2^{C1}, Z_1^{C2}]$)
(5) Let β_2 be coefficient of Z as estimated in previous step
(6) $\hat{\text{ATE}}_{\text{cov}} = \frac{1}{2\alpha-1}(\alpha\beta_1 + (\alpha-1)\beta_2)$

Our method is specified in Algorithm 1. We fit two linear least square estimators; and combine the regression coefficients obtained from them. We use the variable Z as an indicator of treatment allocation. Unlike normal covariate adjustment where outcomes from the same group is chosen; here the model is fit from outcomes of different groups. We fit model with outcomes of treatment 1 from cluster $C1$ and outcomes of treatment 2 from cluster $C2$ and vice versa. The proof of correctness is presented in the Appendix B.

5 EXPERIMENTS & DEMONSTRATION

Description Consider the following scenario. Website 1 and its affiliate website 2, want to do a synchronous change (for example put the same banner on both websites). Ideally with third party cookies this experimented can be conducted in a standard manner. However without such identifier both websites have to individually randomize their treatments.

5.1 Synthetic Data

We first conduct simulation experiments where, by design, all parameters are known and adjustable. We can then quantitatively measure the performance of our method across different ranges of available parameters.

This simulation was conducted by generating 10000 observations, and each experimented was repeated 20 times. of 20 different Poisson variables. Each individual potential outcome was obtained via a noise-corrupted Gaussian distribution. Furthermore each outcome is also influenced by a user specific covariate. Specifically, each potential outcome variable Y_{ij} at a single observation unit is obtained via a linear model from covariates X The true treatment effect in this case is $(1-p)[\mu_{1,0}-\mu_{2,0}]+p(\mu_{1,3}-\mu_{2,4})$. We vary two parameters $\delta_1 = \mu_{1,4}-\mu_{1,3}$ and $\delta_2 = \mu_{2,4}-\mu_{2,3}$

Results We conducted these simulations and measured the error in the estimated ATE versus the true ATE for four methods. These include the standard ATE estimate (uncorrected) , standard ATE with covariate adjustment (uncorrected + adj), our ATE estimate (corrected) and out ATE estimate with covariate adjustment (uncorrected+adj). Due to inherent variability caused by sampling, there will always remain some variability in the estimate. We present in Figure 2 the results from the experiments. The plots depicts both the bias of the estimator and the standard error of the estimator. It is clear that the corrected method provides unbiased estimator of the true ATE while the uncorrected estimator is biased. Furthermore as expected our proposed covariate adjustment method gives an unbiased estimate.

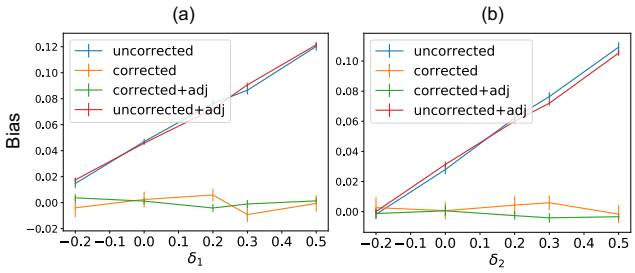


Figure 2: Results on synthetic data. Std error of estimate and bias against variation of interference parameters (a) δ_1 (b) δ_2

5.2 Observational Data

Next we conduct experiments from the experimental logs from Target, an experience cloud product for designing customers' experience. Each entry in the data corresponds to a user visit, and records the users conversion event (which is our targeted binary response) along with the specific experience which the user was served and a variety of covariates such as session duration, browser details, time etc. There are six possible experiences which were regrouped for our purposes into two treatments.

Scenario 1 We split the total time period into two halves, and considered the visitors in the two time periods as visiting two separate websites. This allowed us to simulate the visits on two websites with related treatments. A user who is present in both time periods is a user with exposure to both parts of the treatment, while a user who is present in only one time period corresponds to a user who visits only one website. Splitting the records in a such a way we can create a joint distribution of outcomes and treatments from which samples can be obtained.

Scenario 2 We also ran another experiment where we directly isolated users who got exposure directly to multiple treatments and tried estimated ATE from these experiments as well. Since the outcomes in this case are sampled from historical logs, the effect of interference is fixed and cannot be changed. However one can analyze the effect of changing the fraction of users who receive exposure to multiple treatments.

Results We conducted these simulations and measured the error in the estimated ATE versus the true ATE. We present in Figure 3 the results from the experiments. The plots depicts both the bias of the estimator and the standard error of the estimator. The actual data has less than 0.1% users who received exposure to different treatments, which means the bias in the historical record is minimal. As such we present the results of a simulation done by resampling the records.

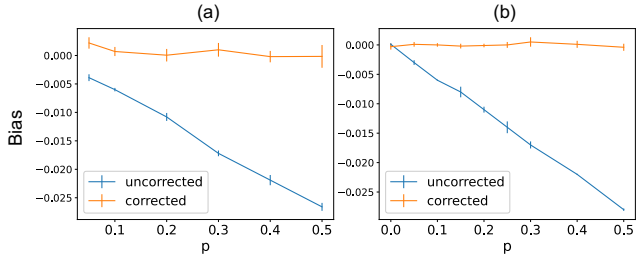


Figure 3: Results on real data. Standard error of estimate and bias against the probability of shared user a) scenario 1 and b) scenario 2

As expected the bias of the standard estimator increases as the probability of user having multiple visits increases. This is in line with the theoretical analysis earlier, which shows that the error is proportional to p (the probability of interfered outcome).

5.3 Demonstration Plan

We develop a demonstration of our technology that shows how a marketer for a brand collaborating on conducting an A/B test jointly with another brand can design and analyse an experiment by only sharing the parameters of the experiment. In Figure 4 the marketer for Brand A specifies the allocation ration of α. The chart shows the estimated treatment effect estimated from the naive

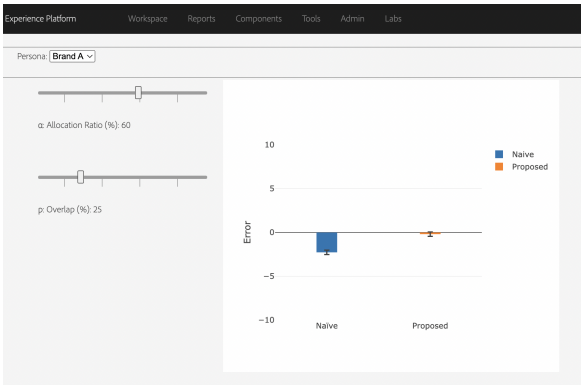


Figure 4: Brand A: With α at 60% and overlap of 25% our method has less error than the naive strategy

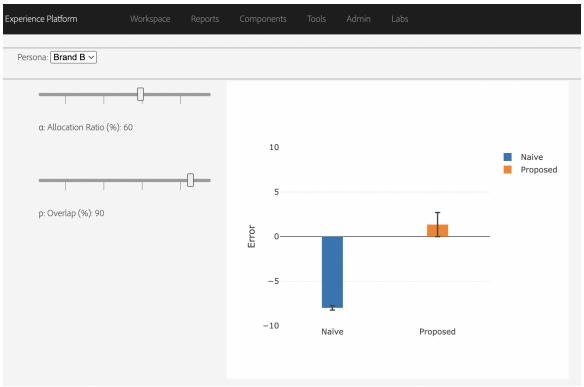


Figure 5: Brand B: With α at 60% and considerable overlap: 90% our proposed method is even better than the naive strategy

difference of treatment differences, and the proposed ATE estimate. Our approach suggests that there is no treatment effect (which is the truth in the simulated scenario), while the naive estimate would suggest a negative treatment effect. Similarly, we show the experience for brand B in Figure 5.

6 CONCLUSION

We have proposed a two-stage experimental study design that two brands can use to jointly test the treatment effects from an experiment. Our approach has no dependence on third-party cookies, and does not require any individual level information to be shared between the two brands. We show that the proposed ATE estimate is unbiased, and give a theoretical formula for the variance of our estimate. We additionally show that we can compute the regression adjusted estimate that can provide additional information on the ATE. Additional work is needed to address multiple treatment arms to control for multiple hypothesis testing. It will also be interesting to explore the effect of identity fragmentation on a single brand's website.

REFERENCES

[1] Matthew Blackwell and Adam N Glynn. 2018. How to make causal inferences with time-series cross-sectional data under selection on observables. *American Political Science Review* 112, 4 (2018), 1067–1082.

[2] Alexander Bleier and Maik Eisenbeiss. 2015. Personalized online advertising effectiveness: The interplay of what, when, and where. *Marketing Science* 34, 5 (2015), 669–688.

[3] Dieter Bohn. 2019. Building a more private web. (2019). https://bit.ly/2ZmoHsY

[4] Patrali Chatterjee, Donna Hoffman, and Thomas Novak. 2003. Modeling the clickstream: Implications for web-based advertising. *Marketing Science* (2003).

[5] Dominic Coey and Michael Bailey. 2016. People and cookies: Imperfect treatment assignment in online experiments. In *Proceedings of WWW*.

[6] Johannes De Smedt, Ewelina Lacka, Spyro Nita, Hans-Helmut Kohls, and Ross Paton. 2021. Session stitching using sequence fingerprinting for web page visits. *Decision Support Systems* 150 (2021), 113579.

[7] William H Greene. 2000. Econometric analysis 4th edition. *International edition, New Jersey: Prentice Hall* (2000), 201–215.

[8] Paul R Hoban and Randolph E Bucklin. 2015. Effects of internet display advertising in the purchase funnel: Model-based insights from a randomized field experiment. *Journal of Marketing Research* 52, 3 (2015), 375–393.

[9] Michael G. Hudgens and M. Elizabeth Halloran. 2008. Toward causal inference with interference. *J. Amer. Statist. Assoc.* 103, 482 (2008), 832–842.

[10] Di Jin, Mark Heimann, Ryan A Rossi, and Danai Koutra. 2019. Node2bits: Compact time-and attribute-aware node representations for user stitching. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 483–506.

[11] Roger Kamena. 2021. Moving towards inferential attribution modelling in a world without third-party cookies. *Applied Marketing Analytics* 7, 2 (2021), 122–130.

[12] Sungchul Kim, Nikhil Kini, Jay Pujara, Eunyee Koh, and Lise Getoor. 2017. Probabilistic visitor stitching on cross-device web logs. In *Proceedings of WWW*.

[13] Selene Leon, Anastasios A Tsiatis, and Marie Davidian. 2003. Semiparametric estimation of treatment effect in a pretest-posttest study. *Biometrics* 59, 4 (2003).

[14] Tesary Lin and Sanjog Misra. 2021. The identity fragmentation bias. (2021).

[15] Puneet Manchanda, Jean-Pierre Dubé, Khim Yong Goh, and Pradeep K Chintagunta. 2006. The effect of banner advertising on internet purchasing. *Journal of Marketing Research* 43, 1 (2006), 98–108.

[16] Klaus M Miller and Bernd Skiera. 2017. Economic loss of cookie lifetime restrictions. *Available at SSRN 3046254* (2017).

[17] Elizabeth L Ogburn, Oleg Sofrygin, Ivan Diaz, and Mark J van der Laan. 2020. Causal inference for social network data. *arXiv preprint arXiv:1705.08527* (2020).

[18] Deborah O'Malley and Joseph Hayon. 2022. The Complete Guide to A/B Testing in a Cookieless World. (2022). https://bit.ly/3fx71cd

[19] Georgia Papadogeorgou, Kosuke Imai, Jason Lyall, and Fan Li. 2020. Causal Inference with Spatio-temporal Data: Estimating the Effects of Airstrikes on Insurgent Violence in Iraq. *arXiv preprint arXiv:2003.13555* (2020).

[20] Judea Pearl. 2000. *Causality: Models, reasoning and inference*. CUP.

[21] Paul R Rosenbaum and Donald B Rubin. 1983. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)* 45, 2 (1983), 212–218.

[22] Oliver J Rutz, Michael Trusov, and Randolph E Bucklin. 2011. Modeling indirect effects of paid search advertising: Which keywords lead to more future visits? *Marketing Science* 30, 4 (2011), 646–665.

[23] Rishiraj Saha Roy, Ritwik Sinha, Niyati Chhaya, and Shiv Saini. 2015. Probabilistic deduplication of anonymous web traffic. In *Proceedings of the 24th International Conference on World Wide Web*. 103–104.

[24] Ritwik Sinha, Ivan Andrus, and Trevor Paulsen. 2020. Attribution IQ: Scalable Game Theoretic Attribution in Web Analytics. In *Proceedings of 29th ACM CIKM*.

[25] Ritwik Sinha, Shiv Saini, and N Anadhavelu. 2014. Estimating the incremental effects of interactions for marketing attribution. In *2014 International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC2014)*. IEEE, 1–6.

[26] Ian Thomas. 2021. Planning for a cookie-less future: How privacy changes will impact marketing, targeting and analytics. *Applied Marketing Analytics* 7 (2021).

[27] Corwin M. Zigler and Georgia Papadogeorgou. 2018. Bipartite Causal Inference with Interference. *arXiv:1807.08660 [stat.ME]* (2018).

A BASIC ESTIMATOR

Our estimator of the treatment effect is given by:

$$\hat{\text{TE}} = \frac{1}{2\alpha - 1}(\alpha \bar{Y}_1^{C1} + (1-\alpha)\bar{Y}_2^{C1} - (1-\alpha)\bar{Y}_1^{C2} - \alpha \bar{Y}_2^{C2}) \tag{2}$$

Claim 1. *$\hat{TE}$ is an unbiased estimate of the treatment effect TE*

Proof. We can see from the earlier discussion that the expected values of the average observed outcomes is given by:

$$\mathbb{E}[\bar{Y}_1^{C1}] = (1-p)\mathbb{E}[Y_{1,0}] + p(\alpha \mathbb{E}[Y_{1,3}] + (1-\alpha)\mathbb{E}[Y_{1,4}])$$
$$\mathbb{E}[\bar{Y}_2^{C1}] = (1-p)\mathbb{E}[Y_{2,0}] + p(\alpha \mathbb{E}[Y_{2,3}] + (1-\alpha)\mathbb{E}[Y_{2,4}])$$
$$\mathbb{E}[\bar{Y}_1^{C2}] = (1-p)\mathbb{E}[Y_{1,0}] + p((1-\alpha)\mathbb{E}[Y_{1,3}] + \alpha \mathbb{E}[Y_{1,4}])$$
$$\mathbb{E}[\bar{Y}_2^{C2}] = (1-p)\mathbb{E}[Y_{2,0}] + p((1-\alpha)\mathbb{E}[Y_{2,3}] + \alpha \mathbb{E}[Y_{2,4}])$$

Now by linearity of expectations:

$$\mathbb{E}[\hat{\text{TE}}] = \frac{1}{2\alpha - 1}(\alpha \mathbb{E}[\bar{Y}_1^{C1}] + (1-\alpha)\mathbb{E}[\bar{Y}_2^{C1}] - (1-\alpha)\mathbb{E}[\bar{Y}_1^{C2}] - \alpha \mathbb{E}[\bar{Y}_2^{C2}])$$

Plugging in the expectations from the earlier equations gives:

$$\mathbb{E}[\hat{\text{TE}}] = (1-p)\mathbb{E}[Y_{1,0} - Y_{2,0}] + p(\mathbb{E}[Y_{1,3} - Y_{2,4}])$$

which is the desired treatment effect . □

The average outcome in each is a combination of the potential outcomes weighted by their relative weights. We have three outcome values, and three effects to estimate: which means that generically these can be uniquely solved.

A.1 Uncertainty Analysis

Since the estimation process is just a linear combination of different average quantities, we can provide an easy upper bound to the variance of our estimator in terms of variance of outcomes. Let V_M and V_m be the maximum and minimum variance among all outcomes i.e $V_M = \max \text{Var}(Y_{1,0}, Y_{2,0}, Y_{1,3}, Y_{1,4}, Y_{2,3}, Y_{2,4})$ and $V_m = \min \text{Var}(Y_{1,0}, Y_{2,0}, Y_{1,3}, Y_{1,4}, Y_{2,3}, Y_{2,4})$. Then, $\frac{V_m}{\sqrt{n}} \leq Var(\bar{Y}_j^{Ci}) \leq \frac{V_M}{\sqrt{n}}$ where n is the number of observations in each group. and

$$\text{Var}(\hat{\text{TE}}) = \frac{2(\alpha^2 + (1-\alpha)^2)}{(2\alpha - 1)^2}\text{Var}(\bar{Y}_j^{Ci})$$

From above expression it is clear that the minimum variance is obtained at $\alpha = 0, 1$. This is not too surprising since that corresponds to an AA vs AB test across the two websites. In such a case we can with certainity find a set of users who exactly receive one treatment (i.e. either 1,3 or 2,4), and the results across the two clusters can be combined to obtain the treatment effect.

An issue however for bounds based on the above variance terms is their validity, since the estimates are only asymptotically normal. However we also note that each average observed effect is an average of outcomes from different users. Hence under independence of each user, Chernoff-Hoeffding bound can be used to provide non-asymptotic intervals for each experiments. Since our final estimator is a linear combination of different average outcome effects, we can trivially obtain valid confidence intervals from the Hoeffding bounds of each individual experiment.

A.2 Estimating conditional effects

Often businesses would be interested in conditional effects as they want to focus more on users who have potentially high conversion rates. From the expression , it is obvious how to obtain conditional average treatment effect (CATE) from the data. As long as conditional outcomes for all four groups can be obtained the earlier expression can be used to obtain the conditional effect by replacing average outcome by conditional average outcomes. These conditionals can be estimated via non-parametric regression, matching or via propensity weighted estimators.

$$\hat{\text{CATE}}|X = \frac{1}{2\alpha - 1}(\alpha \bar{Y}_1^{C1}|X + (1-\alpha)\bar{Y}_2^{C1}|X - (1-\alpha)\bar{Y}_1^{C2}|X - \alpha \bar{Y}_2^{C2}|X) \tag{3}$$

Claim 2. *$\hat{CATE}$ is an unbiased estimate of the conditional average treatment effect CATE*

Proof. The proof of the above statement is analogous to our earlier proof. □

B COVARIATE ADJUSTMENT

Let the numbers of subjects randomized to experimental treatment and control be $n_1 = \sum_{i=1}^{n} Z_i$ and $n_1 = \sum_{i=1}^{n}(1 - Z_i)$, $n = n_0 + n_1$. The sample means of outcome in each group is then given by:

$$\bar{Y}_1 = \frac{1}{n_1}\sum Z_i Y_i$$
$$\bar{Y}_0 = \frac{1}{n_0}\sum (1 - Z_i) Y_i$$
$$\bar{Y} = \frac{1}{n}\sum Y_i$$

Then direct usage of the the OLS formula shows that shows that the least squares estimator for η_Z is

$$\left\{1 - \frac{n^2}{n_0 n_1}\left(n^{-1}d_1\right)^T \sum_{xx}^{-1}\left(n^{-1}d_1\right)\right\}^{-1}\left\{\bar{Y}^{(1)} - \bar{Y}^{(0)} - \frac{n}{n_0 n_1}\sum_{i=1}^{n}\left(Z_i - \bar{Z}\right)\sum_{xY}^{T}\sum_{xX}^{-1} X_i\right\} \tag{4}$$

$$\text{where } d_1 = \sum_{i=1}^{n}\left(Z_i - \bar{Z}\right)X_i,\ \widehat{\sum}_{xy} = n^{-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right), \text{ and}$$

$$\widehat{\sum}_{xx} = n^{-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(X_i - \bar{X}\right)^T$$

Now by Slutsky's theroem, we know that $\hat{\Sigma}_{XY}$ and $\hat{\Sigma}_{XX}$ converge in probability to their expectation counterparts (i.e. the cross-covariance and covariance matrices respectively). Similarly, $n^2/(n_0 n_1) \xrightarrow{p} \{\delta(1-\delta)\}^{-1}$, and $n^{-1}d_1 \xrightarrow{p} 0$.

Hence the first term in (4) asymptotes to 1; while the second term becomes

$$(\bar{Y}_1 - \bar{Y}_0) - \sum (Z_i - \bar{Z})(\sum_{XY}^{T}\sum_{XX}^{-1} X_i)$$

Now since Z is independent of X, the expected value of the second term in the above equation is

$$\mathbb{E}[\sum (Z_i - \bar{Z})\mathbb{E}[(\sum_{XY}^{T}\sum_{XX}^{-1} X_i)] = 0 * \mathbb{E}[(\sum_{XY}^{T}\sum_{XX}^{-1} X_i)] = 0$$

Thus asymptotically the coefficient of Z connverges to the standard estimator. Leon et al. [13] details conditions under which the above estimate has a lower variance. Furthermore well known result in regression theory [7] imply that the above estimator is consistent and asymptotically normal under entirely unrestrictive conditions.

C ASSUMPTIONS

Assumption 1 (Strong ignorability). *Let Z_i^j be a random variable denoting the treatment allocated to user i on site j.*

$$\forall t \quad Z_i^j \perp Y_i^j(t) | \mathbf{X}^i$$

This assumption first explicitly introduced by Rosenbaum and Rubin [21] is standard assumption for causal effect estimation under the potential outcomes approach. This assumption makes the covariates X admissible or deconfounding. Under strong ignorability, treatment effects can be estimated without bias using propensity weighting as shown in Pearl [20].

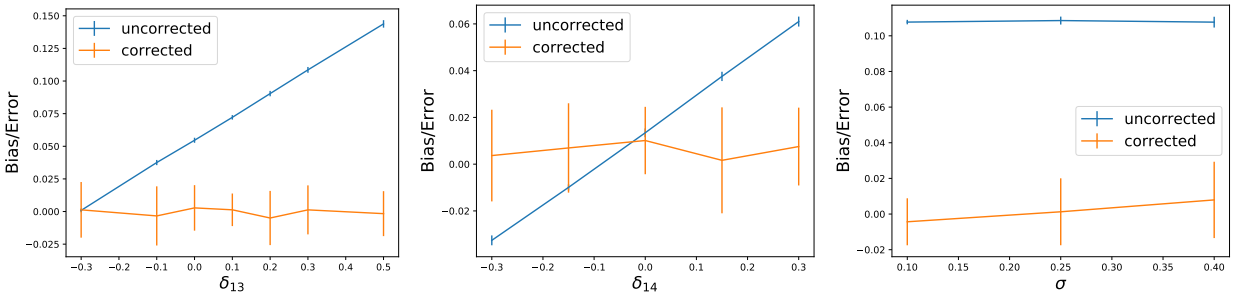
Assumption 2 (Positivity). $0 < P(\mathbf{Z}_i^j = t) < 1$ *for all users i, sites j and treatments t.*

This requirement of positivity (also known as overlapping) ensures that every treatment allocation is possible, or equivalently each unit i has a non-zero probability of being allocated any treatment. Since the different channels are going to independently allocate treatments to any user, this can be easily ensured that there is some randomization at each website.

Assumption 3 (Independent Choice). *If A_i is a random variable denoting the choice of a common user to visit both websites, then $A_i \perp Y_i^j(t) \quad \forall t$.*

While this is implicitly subsumed under the strong ignorability assumption, we make this explicit. The reason for this is that strong ignorability is often used instead of no unobserved confounder. In our scenario no website actually has information about A_i, as they do not know if an individual user will visit the other website, and hence A is unobserved.

D ESTIMATING TREATMENT EFFECT ON ONE CHANNEL



(a) (b) (c)

Figure 6: Results on synthetic data. Std error of estimate and bias against (a) 10 interference (b) 11 interference (c) variance of rv