Designing adaptive experiments for policy learning and inference

Molly Offer-Westort
Assistant Professor, University of Chicago
Political Science Department

Prepared for the
Learning from Interventions Workshop,
Simons Institute for the Theory of Computing
Berkeley, CA
February 16, 2022

Adaptive designs are an exciting frontier in experimental design.

- Allow researchers to select among alternatives, quickly discarding ineffective interventions.
- Assign more observations to the most effective interventions.
- Improve outcomes for respondents as we learn about what works best.

But as a practitioner, how do I go about designing an adaptive experiment that is appropriate for my objectives?

Adaptive designs are an exciting frontier in experimental design.

- Allow researchers to select among alternatives, quickly discarding ineffective interventions.
- Assign more observations to the most effective interventions.
- Improve outcomes for respondents as we learn about what works best.

But as a practitioner, how do I go about designing an adaptive experiment that is appropriate for my objectives?

Motivated by real policymaker concerns!

What this talk is about

- How to design adaptive experiments when you care about inference.
- How to think about writing a design document for an adaptive experiment.
- Some principles for decision making.

- Rapidly growing literature on inference for adaptive experiments.
- Much of the literature on inference for adaptive experiments: if you already have adaptively collected data, what can you do with it to do off-policy evaluation and inference?
- But as a practitioner, how can you set yourself up *at the design phase* to get adaptively collected data that is going to best serve your objective?

Who is this relevant for?

Policy makers who need to make evidence-based decisions when selecting among policy alternatives.

Who is this relevant for?

Policy makers who need to make evidence-based decisions when selecting among policy alternatives.

Policy maker objectives:

- Rigorously establish evidence,

Who is this relevant for?

Policy makers who need to make evidence-based decisions when selecting among policy alternatives.

Policy maker objectives:

- Rigorously establish evidence,
- while maximizing welfare/minimizing harm,

Who is this relevant for?

Policy makers who need to make evidence-based decisions when selecting among policy alternatives.

Policy maker objectives:

- Rigorously establish evidence,
- while maximizing welfare/minimizing harm,
- and learning the most about the policies that are most effective.

Who is this relevant for?

Researchers who want to run a randomized experiment, and who want to learn the “best” operationalization of treatment.

Who is this relevant for?

Researchers who want to run a randomized experiment, and who want to learn the "best" operationalization of treatment.

- Often, when we think about experimental design, we condition on a specific operationalization of treatment.

Who is this relevant for?

Researchers who want to run a randomized experiment, and who want to learn the “best” operationalization of treatment.

- Often, when we think about experimental design, we condition on a specific operationalization of treatment.
- But in practice, there may be a set of treatments whose content is consistent with our research question.

Who is this relevant for?

Researchers who want to run a randomized experiment, and who want to learn the "best" operationalization of treatment.

- Often, when we think about experimental design, we condition on a specific operationalization of treatment.
- But in practice, there may be a set of treatments whose content is consistent with our research question.
- By allowing us to strategically explore alternatives in the treatment space, adaptive designs can allow us to get closer to the theories we're investigating.

The Multi-armed bandit problem

- Algorithms for adaptive experiments navigate the **explore** and **exploit** tradeoff–which is also the problem the policy-maker is facing here.

The Multi-armed bandit problem

- Algorithms for adaptive experiments navigate the **explore** and **exploit** tradeoff–which is also the problem the policy-maker is facing here.
- We can use these methods to run experiments with multiple treatments **more efficiently**.

The Multi-armed bandit problem

- Algorithms for adaptive experiments navigate the **explore** and **exploit** tradeoff–which is also the problem the policy-maker is facing here.
- We can use these methods to run experiments with multiple treatments **more efficiently**.
- But adaptive experiments weren't originally designed with inference and hypothesis testing in mind, so we need to connect to statistical literature for valid inference.

The Multi-armed bandit problem

- Algorithms for adaptive experiments navigate the **explore** and **exploit** tradeoff–which is also the problem the policy-maker is facing here.
- We can use these methods to run experiments with multiple treatments **more efficiently**.
- But adaptive experiments weren't originally designed with inference and hypothesis testing in mind, so we need to connect to statistical literature for valid inference.
- And we may not want to use algorithms out of the box; they are generally designed to maximize welfare/minimize regret, not to provide inferential leverage.

Here are some things we know about inference

Okay, let's say you decide you want to design an adaptive experiment and you care about inference, where should you start?

Bias and adaptively collected data

- If we ignore method of data generation and use sample means, adaptively collected data are going to give us estimates that are systematically negatively biased. (Nie et al., 2017)

Bias and adaptively collected data

- If we ignore method of data generation and use sample means, adaptively collected data are going to give us estimates that are systematically negatively biased. (Nie et al., 2017) Toy Example

We can de-bias with weighting estimators

Inverse Probability Weighting estimation (Horvitz and Thompson, 1952)

For treatment K_i, define $\pi_i(k; S_i) = \Pr[K_i = k | S_i]$.

$$\hat{\mu}_k^{HT} = \frac{1}{N} \sum_{i=1}^{N} Y_i \frac{\mathbb{1}\{K_i = k\}}{\pi_i(k; S_i)}$$

We can de-bias with weighting estimators

Inverse Probability Weighting estimation (Horvitz and Thompson, 1952)

For treatment K_i, define $\pi_i(k; S_i) = \Pr[K_i = k | S_i]$.

$$\hat{\mu}_k^{HT} = \frac{1}{N} \sum_{i=1}^{N} Y_i \frac{\mathbb{1}\{K_i = k\}}{\pi_i(k; S_i)}$$

- (Conditional on algorithm) $\pi_i(k; S_i) > 0, \forall i, k$.

We can de-bias with weighting estimators

Inverse Probability Weighting estimation (Horvitz and Thompson, 1952)

For treatment K_i, define $\pi_i(k; S_i) = \Pr[K_i = k | S_i]$.

$$\hat{\mu}_k^{HT} = \frac{1}{N} \sum_{i=1}^{N} Y_i \frac{\mathbb{1}\{K_i = k\}}{\pi_i(k; S_i)}$$

- (Conditional on algorithm) $\pi_i(k; S_i) > 0, \forall i, k$.
- Finite-n unbiased (Bowden and Trippa, 2017) Finite-n unbiasedness

We can de-bias with weighting estimators

Inverse Probability Weighting estimation (Horvitz and Thompson, 1952)

For treatment K_i, define $\pi_i(k; S_i) = \Pr[K_i = k | S_i]$.

$$\hat{\mu}_k^{HT} = \frac{1}{N} \sum_{i=1}^{N} Y_i \frac{\mathbb{1}\{K_i = k\}}{\pi_i(k; S_i)}$$

- (Conditional on algorithm) $\pi_i(k; S_i) > 0, \forall i, k$.
- Finite-n unbiased (Bowden and Trippa, 2017) Finite-n unbiasedness
- Drawback: if there is low overlap between policy you want to evaluate, and the assigned policy, high variance. One of the tradeoffs we face with adaptively collected data.

Covariate adjustment: doubly robust estimation

We can also consider inverse probability weighting augmented by covariate adjustment.

$$\hat{\mu}_k^{DR} = \frac{1}{N}\sum_{i=1}^{N} \underbrace{\hat{\mathrm{E}}\left[Y_i | K_i = k, X_i\right]}_{\text{estimated outcome model}} + \frac{1}{N}\sum_{i=1}^{N} \underbrace{\frac{\mathbb{1}\{K_i = k\}\left(Y_i - \hat{\mathrm{E}}\left[Y_i | K_i = k, X_i\right]\right)}{\widehat{\pi}_i(k; s)}}_{\text{residual bias correction}}$$

Covariate adjustment: doubly robust estimation

We can also consider inverse probability weighting augmented by covariate adjustment.

$$\hat{\mu}_k^{DR} = \frac{1}{N}\sum_{i=1}^{N} \underbrace{\hat{\mathrm{E}}\left[Y_i | K_i = k, X_i\right]}_{\text{estimated outcome model}} + \frac{1}{N}\sum_{i=1}^{N} \underbrace{\frac{\mathbb{1}\{K_i = k\}\left(Y_i - \hat{\mathrm{E}}\left[Y_i | K_i = k, X_i\right]\right)}{\widehat{\pi}_i(k;s)}}_{\text{residual bias correction}}$$

- We can use regression–or machine-learning methods–for estimation of the outcome model,

Covariate adjustment: doubly robust estimation

We can also consider inverse probability weighting augmented by covariate adjustment.

$$\hat{\mu}_k^{DR} = \frac{1}{N}\sum_{i=1}^{N} \underbrace{\hat{\mathrm{E}}\left[Y_i | K_i = k, X_i\right]}_{\text{estimated outcome model}} + \frac{1}{N}\sum_{i=1}^{N} \underbrace{\frac{\mathbb{1}\{K_i = k\}\left(Y_i - \hat{\mathrm{E}}\left[Y_i | K_i = k, X_i\right]\right)}{\widehat{\pi}_i(k; s)}}_{\text{residual bias correction}}$$

- We can use regression–or machine-learning methods–for estimation of the outcome model,
- or of the weights.

Confidence intervals with adaptively collected data

Even then, with adaptively collected data, we get non-normality in distribution of estimators.

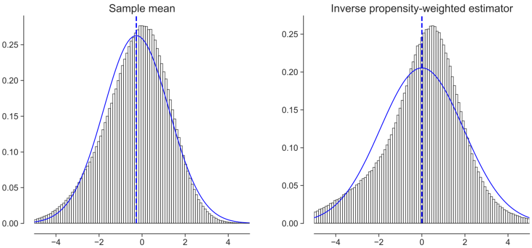


Figure 1: Distribution of the estimates $\widehat{Q}^{\mathrm{AVG}}(1)$ and $\widehat{Q}^{\mathrm{IPW}}(1)$ described in the introduction. The plots depict the distribution of the estimators for $T = 10^6$, scaled by a factor $\sqrt{T}$ for visualization. The distributions are overlaid with the normal curve that matches the first two moments of the distribution, along with a dashed line that denotes the mean. All numbers are aggregated over 1,000,000 replications.

Figure: Hadad et al. (2021), Figure 1

Confidence intervals with adaptively collected data

Some solutions:

- Hadad et al. (2021): adaptively re-weight the terms of the augmented inverse propensity weighting estimator, controlling for contributions of each term to estimator's variance.
  - Requires that the propensity score decay at a slow enough rate.
  - Contextual setting: Zhan et al. (2021)
- Deshpande et al. (2017): augment OLS estimator with $\mathbf{W}-$ "decorrelation" matrix to remove the effects of adaptivity.
- Zhang et al. (2020): batched OLS hypothesis testing; OLS estimates are asymptotically normal *within* batches. Then combine batch-wise statistics and compare to simulated null distribution
  - Requires batched assignment.

Confidence intervals with adaptively collected data

- You still need an appropriate design (sufficient probability floors, batch-wise assignment) for these methods to work.

Confidence intervals: design-based solution

Run an evaluation phase where you use non-adaptive assignment.

Confidence intervals: design-based solution

Run an evaluation phase where you use non-adaptive assignment.

- Often much easier to get right.
  - Often what is done in industry settings.
  - If you're a policy maker and you're trying to make the case for a given policy, this may produce more compelling evidence.
  - If you're evaluating a complicated contextual policy, you may get great returns to efficiency by evaluating on-policy.
    (Adaptively collected data $\rightarrow$ lots of potential for areas with low overlap.)

Confidence intervals: design-based solution

Run an evaluation phase where you use non-adaptive assignment.

- Often much easier to get right.
  - Often what is done in industry settings.
  - If you're a policy maker and you're trying to make the case for a given policy, this may produce more compelling evidence.
  - If you're evaluating a complicated contextual policy, you may get great returns to efficiency by evaluating on-policy. (Adaptively collected data $\rightarrow$ lots of potential for areas with low overlap.)
- Potential concerns.

Confidence intervals: design-based solution

Run an evaluation phase where you use non-adaptive assignment.

- Often much easier to get right.
  - Often what is done in industry settings.
  - If you're a policy maker and you're trying to make the case for a given policy, this may produce more compelling evidence.
  - If you're evaluating a complicated contextual policy, you may get great returns to efficiency by evaluating on-policy. (Adaptively collected data $\rightarrow$ lots of potential for areas with low overlap.)
- Potential concerns.
  - Covariate shift; i.e., different distributions of covariates in the evaluation phase (often easier to address).
  - If you select the wrong arm/policy, you have no chance of getting more data on the true best policy. (Does this always matter? Or is it worth getting good inference on a good enough arm?)

What now?

Ok, so we have some ideas about how to do inference on adaptively collected data, some ideas of what estimators and procedures we might use. What now?

Starting with a design document

Just like any RCT, you should start with a design document. Define how are you going to run your experiment, and how are you going to analyze it.

What hypotheses are you trying to test?

- Adaptively collected data does not generally position us well to make comparisons *across* arms. What if one of our comparison arms only gets a little data?
- If the objective is to make multiple comparisons across arms, non-adaptive RCTs might be better.
- What we can do:
  - Test a hypothesis about each arm independently, (e.g., arm mean is different from zero). Trade off good power, on average, for good arms, with bad power for bad arms.
  - If the control comparison condition is known ex-ante, account for this in your algorithm. (Offer-Westort et al., 2021) algorithm

Why making hypotheses about the "best" arm can be tricky

We might want to compare the best version of treatment to other arms or a control condition. But how do we know which arm is best?

- Suppose you have two fair coins, and you want to estimate the probability of getting heads under the "best" coin.
    1. You flip both coins a number of times.
    2. You take the average portion of times you got heads from each coin.
    3. You take whichever average is higher, and use that as your estimate.
- Conditioning on outcomes is going to give you a biased estimate–even though both coins were fair.

Why making hypotheses about the “best” arm can be tricky

An alternative:

- Repeat steps 1 and 2 above.
  3. Take whichever average is higher, and select that coin as the “best” coin.
  4. *Then* flip that coin a number of times again, and only use these flips to get an estimate of its mean.
- This will produce an unbiased estimate.

Why making hypotheses about the "best" arm can be tricky

Hypothesis wrt the batch-wise "rolling" best arm:

- Let $t \in \{1, \ldots, T\}$ represent batches in the experiment. At each point t in the experiment, we calculate the value of each treatment arm w as the average of scores $\hat{\Gamma}_t(w)$ up to time $t-1$.
- The "best" treatment arm in batch t is the arm with the highest estimate:

$$w^* = \underset{w}{\operatorname{argmax}}\ \hat{\mu}_{t-1}(w)$$

- The score for the best arm in batch t is then the estimate in that period for the selected arm, $\widehat{\Gamma}_t(w^*)$.
- This approach gives us an unbiased estimate of a "best" treatment arm, as in each time period we're conducting learning and evaluation separately. However, note that the underlying treatment arm that is being evaluated is itself random.

Walking through an example

Application: Optimizing messaging in response to vaccine concerns

- Global availability of COVID-19 vaccines is growing.


The New York Times

The Coronavirus Pandemic › | LIVE Covid-19 Updates | Coronavirus

The Next Challenge to Vaccinating Africa: Overcoming Skepticism

Vaccines are finally available in many African countries, but an underfunded public health system has slowed their delivery, and some people there, as well as in South Asia, are wary of taking them.

212 Read in app

By Lynsey Chutel and Max Fisher

Dec. 1, 2021

Application: Optimizing messaging in response to vaccine concerns

- Global availability of COVID-19 vaccines is growing.
- But vaccine hesitancy is limiting government ability to distribute them.


The New York Times

The Coronavirus Pandemic › | LIVE Covid-19 Updates | Coronavirus

The Next Challenge to Vaccinating Africa: Overcoming Skepticism

Vaccines are finally available in many African countries, but an underfunded public health system has slowed their delivery, and some people there, as well as in South Asia, are wary of taking them.

212 | Read in app

By Lynsey Chutel and Max Fisher

Dec. 1, 2021

Application: Optimizing messaging in response to vaccine concerns

Can we train a chatbot to optimize messaging in response to people's stated vaccine concerns?

Application: Optimizing messaging in response to vaccine concerns

Can we train a chatbot to optimize messaging in response to people's stated vaccine concerns?

- Joint work with Leah Rosenzweig (Director, Development Innovation Lab at the Becker Friedman Institute). `https://leahrrosenzweig.com/`
- Supported with work by the Busara Center for Behavioral Economics in Kenya and Nigeria.
- Funded by the Vaccine Confidence Fund.

Treatment space

- We started our project with focus group discussions in Kenya and Nigeria, conducted by the Busara Center for Behavioral Economics, to better understand what people's concerns are, and what information they were interested in to better address these concerns

Treatment space

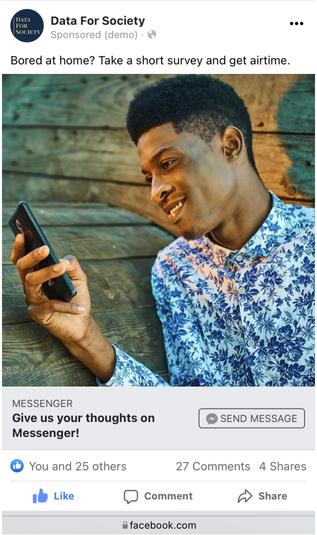
- We started our project with focus group discussions in Kenya and Nigeria, conducted by the Busara Center for Behavioral Economics, to better understand what people's concerns are, and what information they were interested in to better address these concerns
- . . . even if we use great technology for selecting among treatments, if none of them are good, we are not going to get a good outcome.

Treatment space

- We started our project with focus group discussions in Kenya and Nigeria, conducted by the Busara Center for Behavioral Economics, to better understand what people's concerns are, and what information they were interested in to better address these concerns
- . . . even if we use great technology for selecting among treatments, if none of them are good, we are not going to get a good outcome.
- Garbage in $\rightarrow$ garbage out.

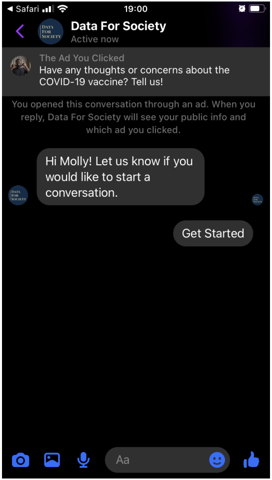
Application: Optimizing messaging in response to vaccine concerns

- Targeted Facebook ads recruit subjects to take a survey, incentivized with phone credit.



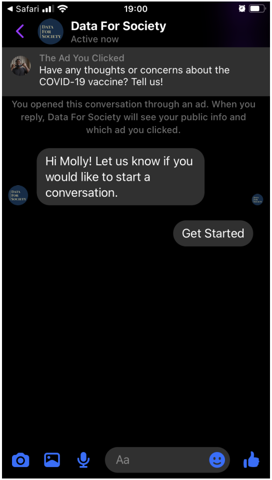
Application: Optimizing messaging in response to vaccine concerns

- Clicking the ad engages users in a Messenger conversation with our chatbot.



Application: Optimizing messaging in response to vaccine concerns

- Clicking the ad engages users in a Messenger conversation with our chatbot.
- We ask people about their positions on the vaccine.



Concern elicitation

- For people who have not yet received the vaccine, we ask about their primary concern or question around the vaccine.

Concern elicitation

- For people who have not yet received the vaccine, we ask about their primary concern or question around the vaccine.
- Concerns are classified into categories, based on concerns identified in previous focus group discussions.

But what messaging is most persuasive?

- We test different messaging interventions related to the concerns that people communicate.

But what messaging is most persuasive?

For example, if someone says their primary concern is **vaccine side effects**, we can send them messaging that addresses:

But what messaging is most persuasive?

For example, if someone says their primary concern is **vaccine side effects**, we can send them messaging that addresses:

- Relatively large risks of COVID-19 infection among unvaccinated individuals.

But what messaging is most persuasive?

For example, if someone says their primary concern is **vaccine side effects**, we can send them messaging that addresses:

- Relatively large risks of COVID-19 infection among unvaccinated individuals.
- Societal benefits of high levels of vaccination rates.

But what messaging is most persuasive?

For example, if someone says their primary concern is **vaccine side effects**, we can send them messaging that addresses:

- Relatively large risks of COVID-19 infection among unvaccinated individuals.
- Societal benefits of high levels of vaccination rates.
- Information that emphasizes that most side effects of the vaccine are mild and short-lived.

But what messaging is most persuasive?

For example, if someone says their primary concern is **vaccine side effects**, we can send them messaging that addresses:

- Relatively large risks of COVID-19 infection among unvaccinated individuals.
- Societal benefits of high levels of vaccination rates.
- Information that emphasizes that most side effects of the vaccine are mild and short-lived.
- Information that debunks common misperceptions about vaccine side effects.

Algorithm selection

What's your objective?

- Best arm selection (+ inference)?
- Minimize regret?
- Contextual policy learning?

Algorithm selection

- Our objective: best arm identification, conditional on concern category.

Algorithm selection

- Our objective: best arm identification, conditional on concern category.
- We use an adaptive algorithm that is optimal for this objective: Top-two Thompson sampling (Russo, 2016).

Algorithm selection

- Our objective: best arm identification, conditional on concern category.
- We use an adaptive algorithm that is optimal for this objective: Top-two Thompson sampling (Russo, 2016).
  - We sample what look like the best two messages (based on posterior probabilities) with equal probability, and all other messages with some minimal probability floor.

Algorithm selection

- Our objective: best arm identification, conditional on concern category.
- We use an adaptive algorithm that is optimal for this objective: Top-two Thompson sampling (Russo, 2016).
  - We sample what look like the best two messages (based on posterior probabilities) with equal probability, and all other messages with some minimal probability floor.
  - The algorithm learns which messages are best throughout the experiment, and updates based on new data.

Sample size determination

- Suppose you care most about inference wrt the best arm, but you don't know ex ante which arm is best, or how much sample you're going to assign to it, what probability weights will be. How do you do a power calculation?
- If the algorithm is able to identify the best arm very quickly, you will get a lot of data, assign treatment to the best arm with high probability, and will get good estimates.
- What is a feasible worst case version of the world?

Sample size determination

- Our objective: best arm identification, conditional on concern category.

Sample size determination

- Our objective: best arm identification, conditional on concern category.
- Outcome is binary response. Start with a mean of 0.5 to maximize variance.

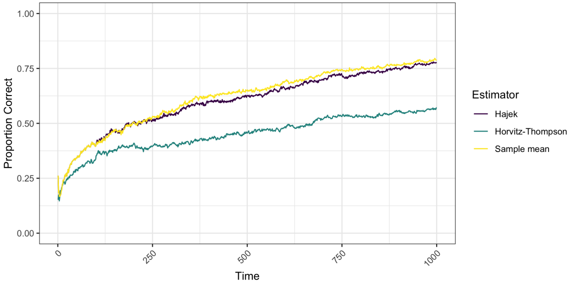
Sample size determination

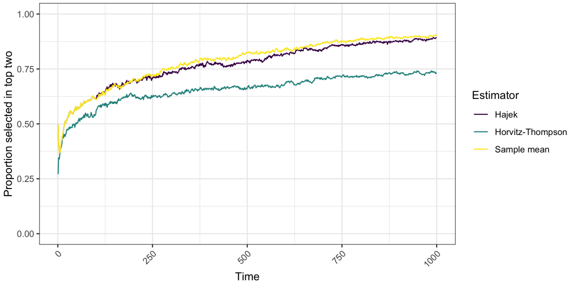
- Our objective: best arm identification, conditional on concern category.
- Outcome is binary response. Start with a mean of 0.5 to maximize variance.
- What if the best arm is only 0.01 points better than the second best arm?

Sample size determination

- Our objective: best arm identification, conditional on concern category.
- Outcome is binary response. Start with a mean of 0.5 to maximize variance.
- What if the best arm is only 0.01 points better than the second best arm?
- What if the second best arm is only 0.01 points better than remaining arms?

Sample size determination

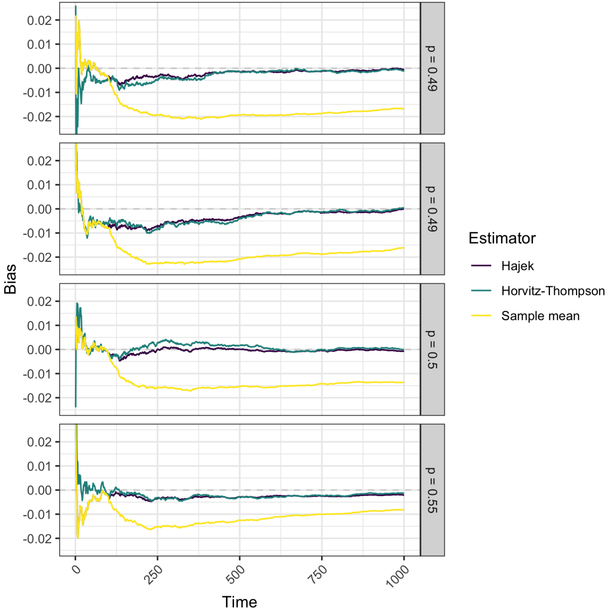




Sample size determination

- Determining feasible worst case version of the world is harder with contextual cases.
- Consider: what if the optimal contextual policy is only X standard deviations better than the best fixed policy?

Estimating procedures



Balancing weights

- Sample mean does pretty well for best arm selection, but is systematically biased.
- It can happen that our algorithm selects a “best” arm that is not consistent with what we get if we used IPW estimates.
- To resolve this inconsistency, we can incorporate balancing weights into the algorithm, per Dimakopoulou et al. (2019).

Other design parameters

Batched vs. online?

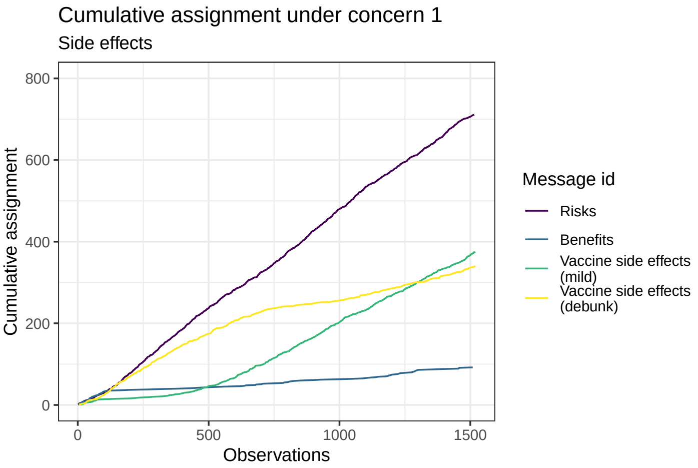
- How large is your first batch? (Even if you're doing online, still a good idea to do uniform assignment in first batch, for stability of algorithm). A big first batch can help with more stability of the algorithm–esp. important if using IPW estimates.
- If batched, how frequently do you update? After the first batch, generally more is better, but often by not a lot. This is partially an implementation consideration, how often can you update realistically?

Other design parameters

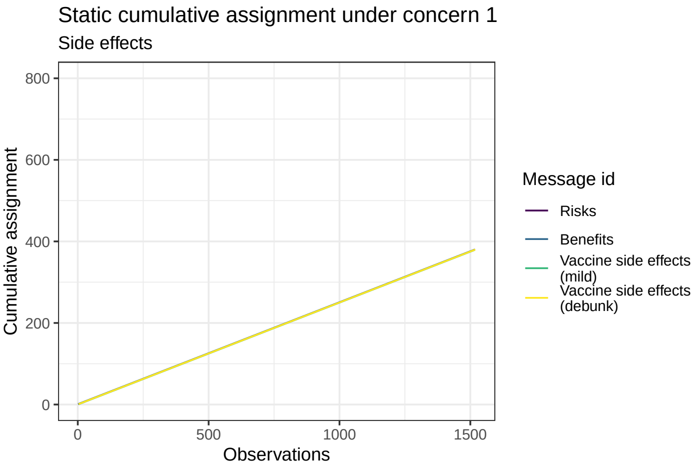
Probability floors

- Ensure all arms continue to be sampled with some probability.
- Increases regret.
- Contributes to behavior of estimators, see e.g., Hadad et al. (2021).

Adaptive assignment



[Hypothetical] Static assignment



Results

- For the best arm, we assign nearly 700 observations; compared to $\approx$ 370 under a conventional static experiment.

Results

- For the best arm, we assign nearly 700 observations; compared to $\approx$ 370 under a conventional static experiment.
- During the adaptive portion of this experiment, we learn whether respondents say that messaging addressed their concerns.

Results

- For the best arm, we assign nearly 700 observations; compared to $\approx$ 370 under a conventional static experiment.
- During the adaptive portion of this experiment, we learn whether respondents say that messaging addressed their concerns.
- Our estimate of response under the best arm is much more precise–standard errors are 72% of the size they would be under a static design.

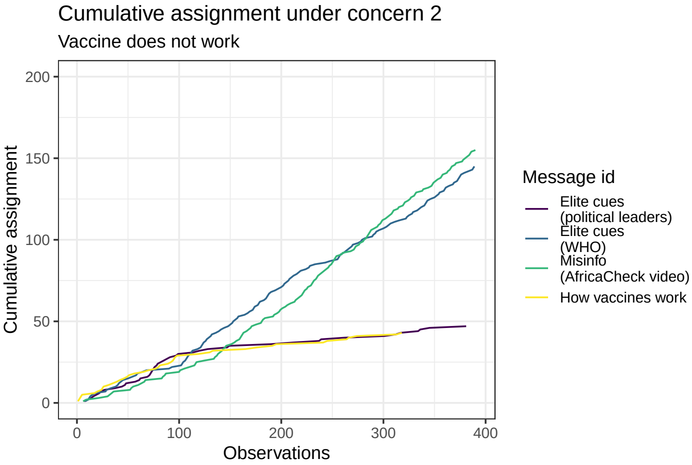
Other concerns

We have seven pre-determined concern categories:

1. I'm worried about the side effects of the vaccine.
2. I don't know if the vaccine works.
3. I don't think COVID-19 is real.
4. I don't need the vaccine because I am protected by God.
5. I don't trust healthcare workers.
6. I don't trust the government.
7. I am hearing different things about the vaccine, and I'm not sure what to believe.

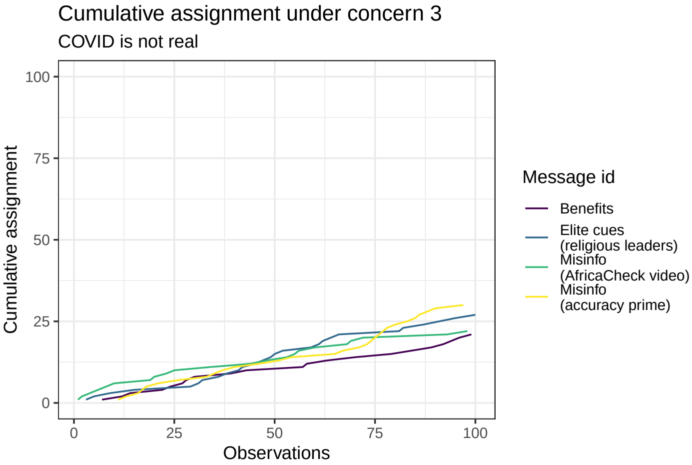
Other concerns

How effective the algorithm is depends on how much data we have for each concern.



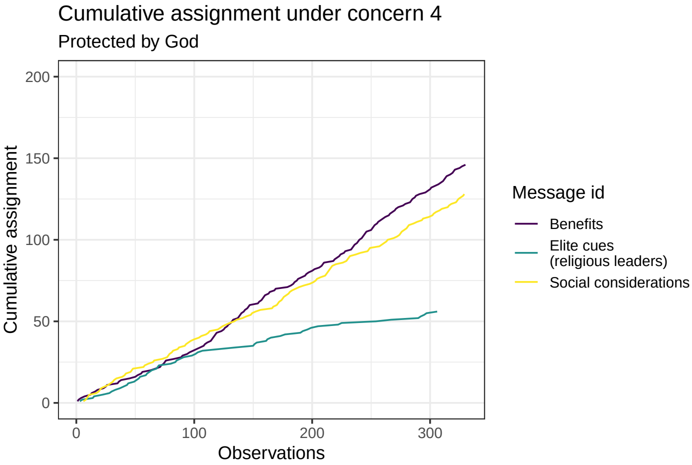
Other concerns

How effective the algorithm is depends on how much data we have for each concern.



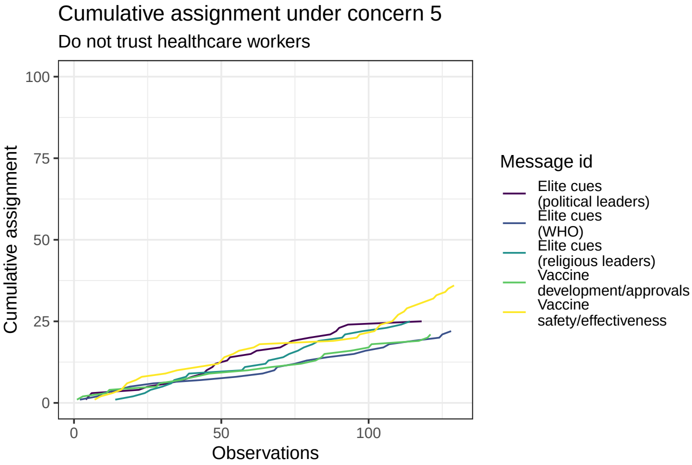
Other concerns

How effective the algorithm is depends on how much data we have for each concern.



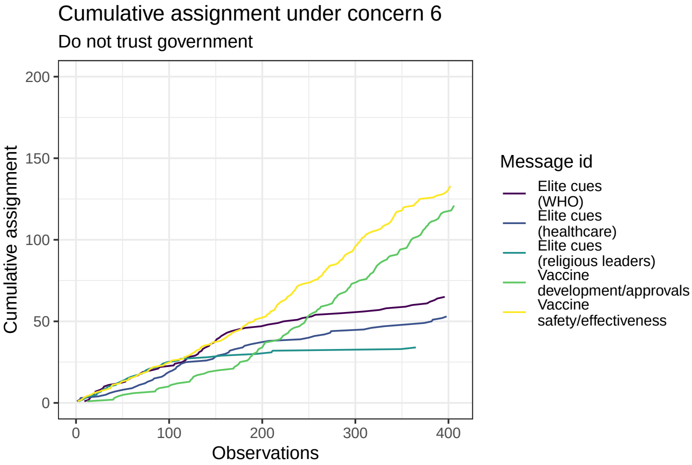
Other concerns

How effective the algorithm is depends on how much data we have for each concern.



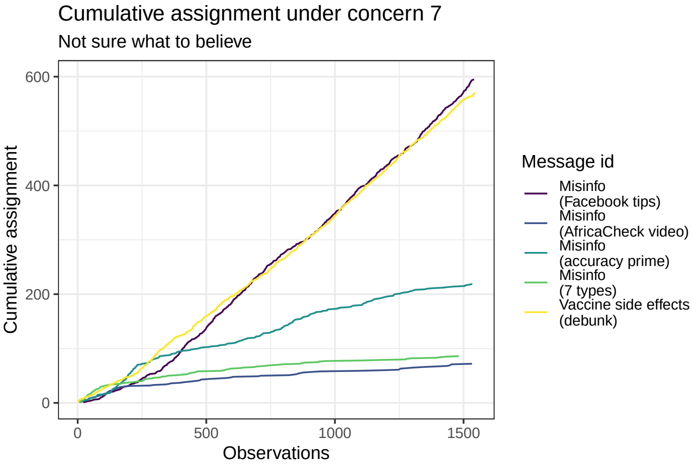
Other concerns

How effective the algorithm is depends on how much data we have for each concern.



Other concerns

How effective the algorithm is depends on how much data we have for each concern.



Next steps

- Once we learn which messages are optimal, we'll run our chatbot at scale.

Next steps

- Once we learn which messages are optimal, we'll run our chatbot at scale.
- The outcome we care about is whether people report that they intend to get vaccinated.

Next steps

- Once we learn which messages are optimal, we'll run our chatbot at scale.
- The outcome we care about is whether people report that they intend to get vaccinated. And, three weeks later, if they were more likely to have actually gotten vaccinated.

Next steps

- Once we learn which messages are optimal, we'll run our chatbot at scale.
- The outcome we care about is whether people report that they intend to get vaccinated. And, three weeks later, if they were more likely to have actually gotten vaccinated.
- We will evaluate how effective our concern-eliciting chatbot is, compared to a control condition.

Heterogeneous treatment effects: Contextual bandits

Context matters: We think that people with different political preferences *also* may respond differently to interventions.

Heterogeneous treatment effects: Contextual bandits

Context matters: We think that people with different political preferences *also* may respond differently to interventions.

- We could run separate adaptive procedures based on political affiliation.
- Facilitates separate learning processes.

Heterogeneous treatment effects: Contextual bandits

Context matters: We think that people with different political preferences *also* may respond differently to interventions.

- We could run separate adaptive procedures based on political affiliation.
- Facilitates separate learning processes.

What if you have more complex covariates. . . ?

Heterogeneous treatment effects: Contextual bandits

Context matters: We think that people with different political preferences *also* may respond differently to interventions.

- We could run separate adaptive procedures based on political affiliation.
- Facilitates separate learning processes.

What if you have more complex covariates. . . ?

- Instead of thinking of best fixed treatments, we would like to find the best **personalized policy**.

Heterogeneous treatment effects: Contextual bandits

Context matters: We think that people with different political preferences *also* may respond differently to interventions.

- We could run separate adaptive procedures based on political affiliation.
- Facilitates separate learning processes.

What if you have more complex covariates. . . ?

- Instead of thinking of best fixed treatments, we would like to find the best **personalized policy**.
- We can then think about testing hypotheses about best personalized policies vs. best fixed.
- Challenges: if dimension of covariates is large, may need more data to learn a good policy.

Take-aways

- Design documents are even more important for adaptive studies, as compared to non-adaptive experiments.
- Determine your objectives. What hypotheses will you test?
- Be explicit about your algorithm. Include your code if possible.
- Conduct sample size determination via simulations. When running simulations, consider feasible worst case scenarios.

Shiny app

https://mollyow.shinyapps.io/adaptive/

Thank you to collaborators

Leah Rosenzweig
Susan Athey
P M Aronow
Donald P. Green
Alexander Coppock

mollyow@uchicago.edu
mollyow.github.io

References

Bowden, J. and Trippa, L. (2017). Unbiased estimation for response adaptive clinical trials. Statistical methods in medical research, 26(5):2376–2388.

Deshpande, Y., Mackey, L., Syrgkanis, V., and Taddy, M. (2017). Accurate inference for adaptive linear models. arXiv preprint arXiv:1712.06695.

Dimakopoulou, M., Athey, S., and Imbens, G. (2017). Estimation considerations in contextual bandits. arXiv preprint arXiv:1711.07077.

Dimakopoulou, M., Zhou, Z., Athey, S., and Imbens, G. (2019). Balanced linear contextual bandits. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 3445–3453.

Gosnell, H. F. (1927). Getting out the vote. University of Chicago Press Chicago.

Hadad, V., Hirshberg, D. A., Zhan, R., Wager, S., and Athey, S. (2021). Confidence intervals for policy evaluation in adaptive experiments. Proceedings of the National Academy of Sciences, 118(15).

Harris, J. P. (1927). Getting out the vote. by harold f. gosnell (chicago: University of chicago press. 1927. pp. xi, 128.). American Political Science Review, 21(4):902–903.

Horvitz, D. and Thompson, D. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47(260):663–685.

Nie, X., Tian, X., Taylor, J., and Zou, J. (2017). Why Adaptively Collected Data Have Negative Bias and How to Correct for It. arXiv e-prints arXiv:1708.01977.

Offer-Westort, M., Coppock, A., and Green, D. P. (2021). Adaptive experimental design: Prospects and applications in political science. American Journal of Political Science, 65(4):826–844.

Russo, D. (2016). Simple bayesian algorithms for best arm identification. In Conference on Learning Theory, pages 1417–1418. PMLR.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika, 25(3/4):285–294.

Zhan, R., Hadad, V., Hirshberg, D. A., and Athey, S. (2021). Off-policy evaluation via adaptive weighting with data from contextual bandits. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pages 2125–2135.

Zhang, K. W., Janson, L., and Murphy, S. A. (2020). Inference for batched bandits. arXiv preprint arXiv:2002.03217.

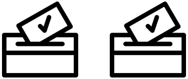
Appendix

Bias in naive estimation w/ adaptive design

Treatments	 A	 B
(Unknown) mean response	0.5	0.5

Toy example

Bias in naive estimation w/ adaptive design



Treatments	A	B
(Unknown) mean response	0.5	0.5

Toy example

- **Period 1:** Assign **one subject each** to treatment A, treatment B.

Bias in naive estimation w/ adaptive design

Treatments	 **A**	 **B**
(Unknown) mean response	0.5	0.5

Toy example

- **Period 1:** Assign **one subject each** to treatment A, treatment B.

$\Rightarrow$ Observe data.

Bias in naive estimation w/ adaptive design

Treatments	 **A**	 **B**
(Unknown) mean response	0.5	0.5

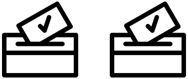
Toy example

- **Period 1:** Assign **one subject each** to treatment A, treatment B.

⇒ Observe data.

- **Period 2:** Assign treatment probabilistically to **one subject** using Thompson sampling.

Bias in naive estimation w/ adaptive design



Treatments	**A**	**B**
(Unknown) mean response	0.5	0.5

Toy example

- **Period 1:** Assign **one subject each** to treatment A, treatment B.

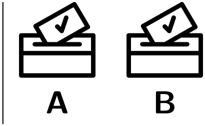
⇒ Observe data.

- **Period 2:** Assign treatment probabilistically to **one subject** using Thompson sampling.

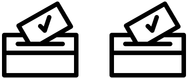
⇒ Three subjects total, two periods. Estimate sample means.

Bias in naive estimation w/ adaptive design: one realized experiment

Treatments



Bias in naive estimation w/ adaptive design: one realized experiment



Treatments	**A**	**B**
Period 1		
- sample	A	B

Bias in naive estimation w/ adaptive design: one realized experiment

Treatments	**A**	**B**
Period 1		
- sample	A	B
- outcome	1	0

Bias in naive estimation w/ adaptive design: one realized experiment

Treatments	 **A**	 **B**
Period 1		
- sample	A	B
- outcome	1	0
Sampling probability	5/6	1/6

Bias in naive estimation w/ adaptive design: one realized experiment

Treatments	 **A**	 **B**
Period 1		
- sample	A	B
- outcome	1	0
Sampling probability	5/6	1/6
Period 2		
- sample	A	-

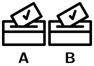
Bias in naive estimation w/ adaptive design: one realized experiment

Treatments	 **A**	 **B**
Period 1		
- sample	A	B
- outcome	1	0
Sampling probability	5/6	1/6
Period 2		
- sample	A	-
- outcome	0	-

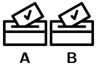
Bias in naive estimation w/ adaptive design: one realized experiment

Treatments	 **A**	 **B**
Period 1		
- sample	A	B
- outcome	1	0
Sampling probability	5/6	1/6
Period 2		
- sample	A	-
- outcome	0	-
Sample mean	0.5	0

Bias in naive estimation w/ adaptive design: all paths



Bias in naive estimation w/ adaptive design: all paths



Period 1

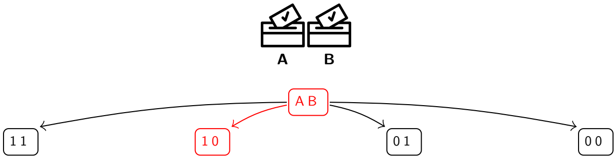
- sample

A B

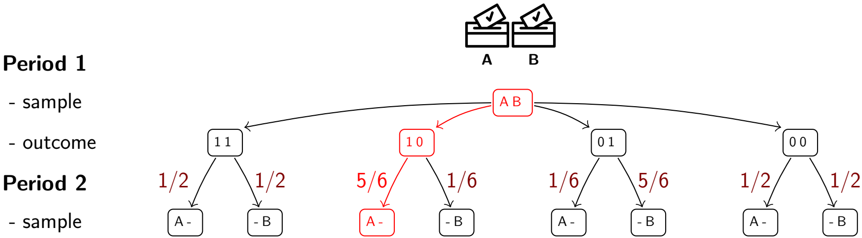
Bias in naive estimation w/ adaptive design: all paths

Period 1

- sample
- outcome



Bias in naive estimation w/ adaptive design: all paths



Bias in naive estimation w/ adaptive design: all paths

Period 1

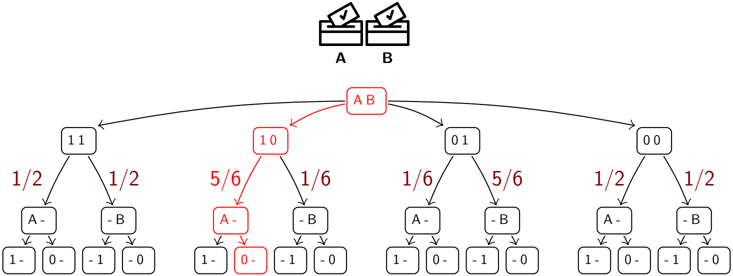
- sample

- outcome

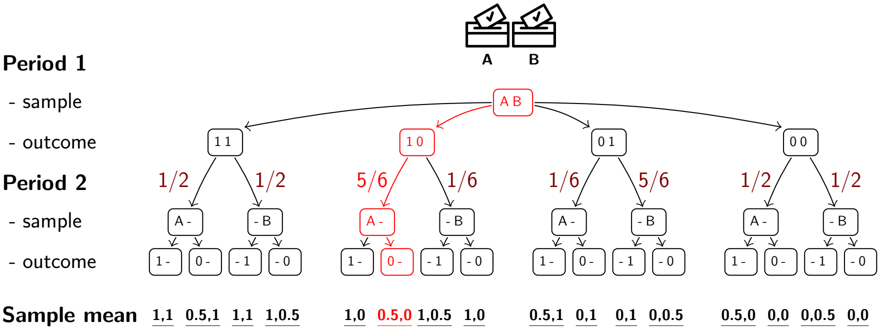
Period 2

- sample

- outcome

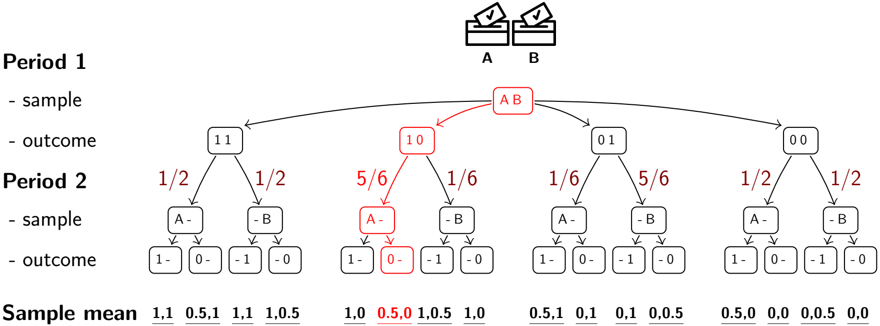


Bias in naive estimation w/ adaptive design: all paths



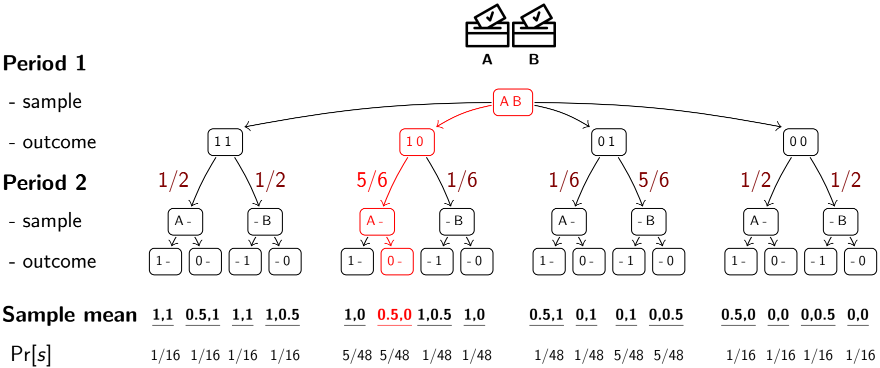
Bias in naive estimation w/ adaptive design: all paths

After observing three outcomes, bias is negative.



Bias in naive estimation w/ adaptive design: all paths

After observing three outcomes, bias is negative.



$$\mathrm{E}\left[\widehat{\mu}_k\right] = \sum_s \widehat{\mu}_k | s \cdot \Pr[s]$$

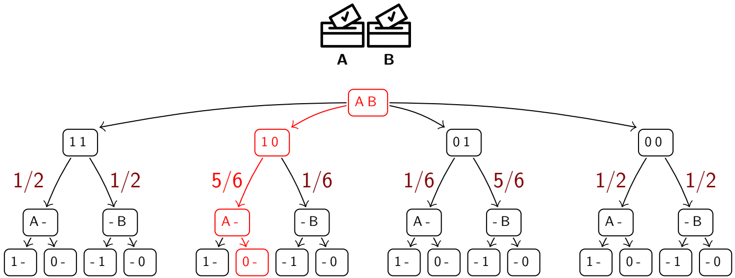
$$\approx 0.458$$

Bias in naive estimation w/ adaptive design: all paths

After observing three outcomes, bias is negative.

Period 1
- sample
- outcome

Period 2
- sample
- outcome



Sample mean	1,1	0.5,1	1,1	1,0.5	1,0	0.5,0	1,0.5	1,0	0.5,1	0,1	0,1	0,0.5	0.5,0	0,0	0,0.5	0,0
Pr[s]	1/16	1/16	1/16	1/16	5/48	5/48	1/48	1/48	1/48	1/48	5/48	5/48	1/16	1/16	1/16	1/16

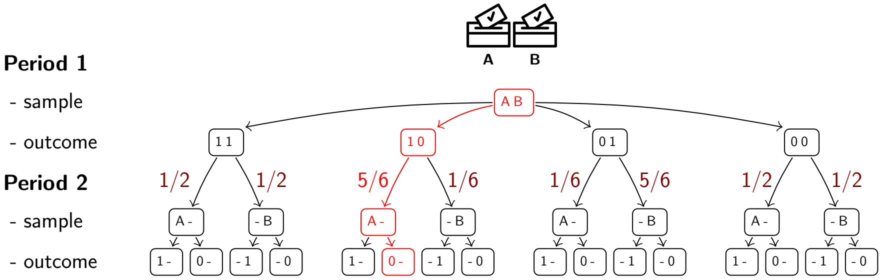
$$\mathrm{E}\left[\widehat{\mu}_k\right] = \sum_s \widehat{\mu}_k | s \cdot \Pr[s]$$

$$\approx 0.458$$

(S_i is a state, defined by the history of treatment and outcomes.)

Bias in naive estimation w/ adaptive design: all paths

After observing three outcomes, bias is negative.



Sample mean	**1,1**	**0.5,1**	**1,1**	**1,0.5**	**1,0**	**0.5,0**	**1,0.5**	**1,0**	**0.5,1**	**0,1**	**0,1**	**0,0.5**	**0.5,0**	**0,0**	**0,0.5**	**0,0**
Pr[s]	1/16	1/16	1/16	1/16	5/48	5/48	1/48	1/48	1/48	1/48	5/48	5/48	1/16	1/16	1/16	1/16

Why? Lower probability of sampling from an arm if it performs poorly in the first period.

Back to main slides

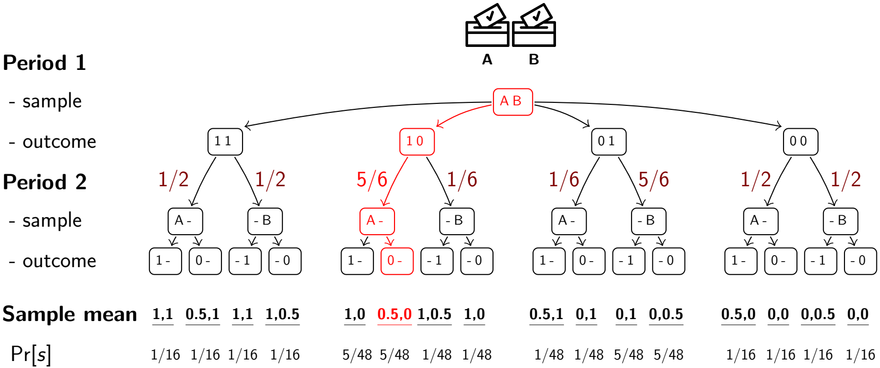
Estimation with Inverse Probability Weighting

$$\hat{\mu}_k^{HT} = \frac{1}{N} \sum_{i=1}^{N} Y_i \frac{\mathbb{1}\{K_i = k\}}{\pi_i(k; S_i)}$$

In our example...

$$\begin{aligned}\hat{\mu}_A^{HT} &= \frac{1}{3} \left[1 \times \frac{\mathbb{1}\{K_i = A\}}{1/2} + 0 \times \frac{\mathbb{1}\{K_i = A\}}{5/6} \right] \\ &= \frac{2}{3}\end{aligned}$$

De-biasing estimation: all paths



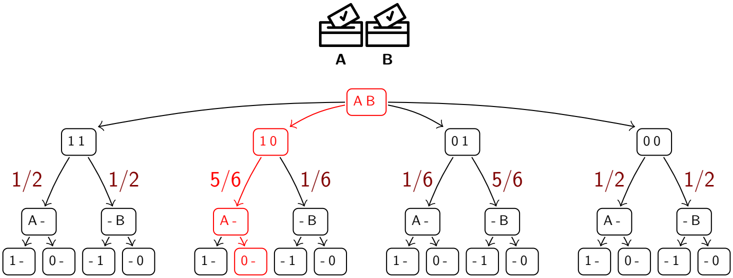
De-biasing estimation: all paths

Period 1

- sample
- outcome

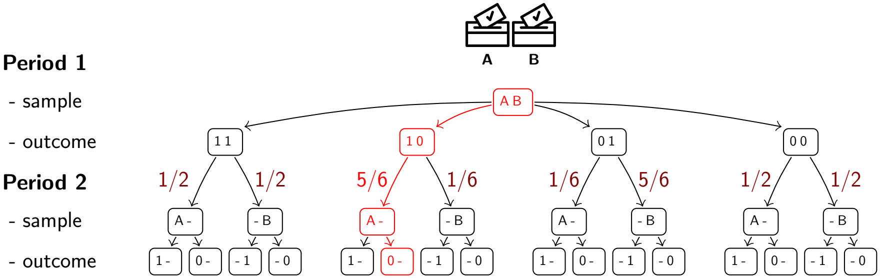
Period 2

- sample
- outcome



HT estimate	$\frac{4}{3},\frac{2}{3}$	$\frac{2}{3},\frac{2}{3}$	$\frac{2}{3},\frac{4}{3}$	$\frac{2}{3},\frac{2}{3}$	$\frac{16}{15}$,0	$\frac{2}{3}$,0	$\frac{2}{3}$,2	$\frac{2}{3}$,0	2,$\frac{2}{3}$	0,$\frac{2}{3}$	0,$\frac{16}{15}$	0,$\frac{2}{3}$	$\frac{2}{3}$,0	0,0	0,$\frac{2}{3}$	0,0
$\Pr[s]$	1/16	1/16	1/16	1/16	5/48	5/48	1/48	1/48	1/48	1/48	5/48	5/48	1/16	1/16	1/16	1/16

De-biasing estimation: all paths



HT estimate	$\frac{4}{3},\frac{2}{3}$	$\frac{2}{3},\frac{2}{3}$	$\frac{2}{3},\frac{4}{3}$	$\frac{2}{3},\frac{2}{3}$	$\frac{16}{15},0$	$\frac{2}{3},0$	$\frac{2}{3},2$	$\frac{2}{3},0$	$2,\frac{2}{3}$	$0,\frac{2}{3}$	$0,\frac{16}{15}$	$0,\frac{2}{3}$	$\frac{2}{3},0$	$0,0$	$0,\frac{2}{3}$	$0,0$
$\Pr[s]$	1/16	1/16	1/16	1/16	5/48	5/48	1/48	1/48	1/48	1/48	5/48	5/48	1/16	1/16	1/16	1/16

$$\mathrm{E}\left[\widehat{\mu}_k^{HT}\right] = \sum_s \widehat{\mu}_k^{HT}|s \cdot \Pr[s]$$

$$= 0.5$$

Back to main slides

HT unbiased

Thompson sampling algorithm

We suppose that K arms have unknown success rates $\theta_1, \ldots, \theta_K$, following their respective Bernoulli distributions, with likelihoods

$$f_{X_1|\Theta_1}(x_1|\theta_1), \ldots, f_{X_K|\Theta_K}(x_K|\theta_K).$$

Posteriors follow Beta distributions with parameters $\alpha_{k,t}, \beta_{k,t}$.

Algorithm 1: Batch-wise Thompson sampling

1: Initialize priors such that $(\alpha_{k,1} = 1, \beta_{k,1} = 1)$ for $k = 1, \ldots, K$.

For periods $t = 1, \ldots, T$:

2: Calculate $p_{k,t} = P\left[\Theta_k = \max_k\{\Theta_1, \ldots, \Theta_K\}|(\alpha_{1,t}, \beta_{1,t}), \ldots, (\alpha_{K,t}, \beta_{K,t})\right]$ for $k = 1, \ldots, K$.*

3: Sample n observations, assigning treatment with probabilities $(p_{1,t}, \ldots, p_{K,t})$.

4: Update posteriors, for $k = 1, \ldots, K$:

$$\alpha_{k,t+1} = \alpha_{k,t} + \#\text{ successes observed for arm } k \text{ in period } t,$$
$$\beta_{k,t+1} = \beta_{k,t} + \#\text{ failures observed for arm } k \text{ in period } t.$$

Control augmented Thompson sampling algorithm

Algorithm 2: Control-augmented Thompson sampling

1: Let C index the control arm. Initialize priors such that $(\alpha_{k,1} = 1, \beta_{k,1} = 1)$ for $k \neq C$.
For periods $t = 1, \ldots, T$:

2: Calculate $p_{k,t}$ as above in step 2, excluding C.

3: Retrieve the "current best arm," and calculate the difference between the cumulative sample assigned to that arm and the control arm:

$$b = \arg\max_k p_{k,t},$$
$$d = n_{b,t} - n_{C,t}.$$

4: Calculate the proportion of the next batch needed for the control to match the cumulative sample of the "best" arm, up to a maximum of Z_t of the batch, where $Z_t \in (0,1)$ may be fixed, or may be data adaptive:

$$q = \min(\max(d/n, 0), Z_t).$$

5: The probability of assignment to the control condition is a combination of the allocation to match the cumulative sample of the control to the current best arm, and R_t of the remaining probability, for $R_t \in (0,1)$. This value may be fixed, or may also be data adaptive:

$$\tilde{p}_{C,t} = q + R_t * (1 - q).$$

6: The treatment arms are assigned according to their posterior probabilities, scaled to the remaining sampling probability, for $k \neq C$:

$$\tilde{p}_{k,t} = p_{k,t} * (1 - R_t) * (1 - q).$$

7: Sample n observations, assigning treatment with probabilities $(\tilde{p}_{1,t}, \ldots, \tilde{p}_{C,t}, \ldots, \tilde{p}_{K,t})$.

8: Update posteriors, for $k \neq C$:

$$\alpha_{k,t+1} = \alpha_{k,t} + \# \text{ successes observed for arm } k \text{ in period } t,$$
$$\beta_{k,t+1} = \beta_{k,t} + \# \text{ failures observed for arm } k \text{ in period } t.$$

back to hypotheses

Finite-n unbiasedness of the Horvitz-Thompson estimator

Under the potential outcomes framework:

$$\mathrm{E}\left[\hat{\mu}_k^{HT}\right] = \frac{1}{N}\sum_{i=1}^{N} Y_i(k).$$

We require only independence of potential outcomes and treatment conditional on history, $Y_i(k) \perp\!\!\!\perp K_i|S_i$, which is given by the experimental design.

$$\mathrm{E}\left[\hat{\mu}_k^{HT}\right] = \mathrm{E}\left[\frac{1}{N}\sum_{i=1}^{N} Y_i \frac{\mathbb{1}\{K_i = k\}}{\pi_i(k;S_i)}\right]$$
$$= \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}\left[Y_i \frac{\mathbb{1}\{K_i = k\}}{\pi_i(k;S_i)}\right].$$

Considering the i^{th} unit, by the Law of Iterated Expectations,

$$\mathrm{E}\left[Y_i \frac{\mathbb{1}\{K_i = k\}}{\pi_i(k;S_i)}\right] = \mathrm{E}\left[\mathrm{E}\left[Y_i \frac{\mathbb{1}\{K_i = k\}}{\pi_i(k;S_i)}\middle| S_i\right]\right]$$

Taking the interior term, $\mathrm{E}\left[Y_i \frac{\mathbb{1}\{K_i=k\}}{\pi_i(k;S_i)}\middle| S_i\right]$, by definition,

$$= \mathrm{E}\left[\frac{Y_i \mathbb{1}\{K_i = k\}}{\Pr[K_i = k|S_i]}\middle| S_i\right]$$

By the potential outcomes model,

$$= \mathrm{E}\left[Y_i(k) \times \frac{\mathbb{1}\{K_i = k\}}{\Pr[K_i = k|S_i]}\middle| S_i\right]$$

And because $Y_i \perp\!\!\!\perp K_i|S_i$,

$$= \mathrm{E}\left[Y_i(k)|S_i\right] \times \mathrm{E}\left[\frac{\mathbb{1}\{K_i = k\}}{\Pr[K_i = k|S_i]}\middle| S_i\right]$$
$$= \mathrm{E}\left[Y_i(k)|S_i\right] \times \frac{\mathrm{E}\left[\mathbb{1}\{K_i = k\}|S_i\right]}{\Pr[K_i = k|S_i]}$$
$$= \mathrm{E}\left[Y_i(k)|S_i\right] \times \frac{\Pr[K_i = k|S_i]}{\Pr[K_i = k|S_i]}$$
$$= \mathrm{E}\left[Y_i(k)|S_i\right].$$

Then returning to the Law of Iterated Expectations from above,

$$\mathrm{E}\left[Y_i \frac{\mathbb{1}\{K_i = k\}}{\pi_i(k;S_i)}\right] = \mathrm{E}\left[\mathrm{E}\left[Y_i(k)|S_i\right]\right] = \mathrm{E}\left[Y_i(k)\right].$$

HT estimator

Equivalence to sample mean in static trials

For static designs with complete random assignment, the Horvitz-Thompson estimate is equivalent to the sample mean, as $\pi_i(k) = \frac{1}{K}$ and there are N/K subjects assigned each treatment.

$$\begin{aligned}\hat{\mu}_k^{HT} &= \frac{1}{N}\sum_{i=1}^{N} Y_i \frac{\mathbb{1}\{K_i = k\}}{\pi_i(k)} \\ &= \frac{1}{N/K}\sum_{i=1}^{N} Y_i \mathbb{1}\{K_i = k\} \\ &= \frac{1}{N/K}\sum_{i:K_i=k} Y_i\end{aligned}$$