Adaptive Experiments for Policy Choice

Maximilian Kasy Anja Sautmann

December 7, 2018

Introduction

- Consider an NGO that has the goal of encouraging "kangaroo care" for prematurely born babies.
- There are numerous implementation choices:
  - Incentives for health-care providers,
  - educating mothers, ...
- We argue:
  - NGO should run an experiment in multiple waves.
  - Initially, try many different variants.
  - Later, focus the experiment on the best performing options.
  - Once the experiment is concluded, recommend the best performing option.
- Principled approach for pilot studies, or "tinkering."
- In the spirit of "the economist as plumber" (Duflo, 2017).

Introduction

- Our setting:
  - Multiple waves.
  - Objective:
    1. After the experiment pick a policy
    2. to maximize social welfare.
  - How to design experiments for this objective?
- Contrast with canonical field experiments:
  - One wave.
  - Objectives:
    1. Estimate average treatment effect.
    2. Test whether it equals 0.
  - Design recommendations:
    1. Same number of observations for each treatment.
    2. If possible stratify.
    3. Choose sample size based on power calculations.

Introduction

Preview of findings

- The distinction matters:
  - Optimal designs look qualitatively different for different objective functions.
  - Adaptive designs for policy choice improve welfare.
- Implementation:
  - Optimal designs are feasible but computationally challenging.
  - Good and easily computed approximations are available.
- Features of optimal designs:
  - Adapt to the outcomes of previous waves.
  - Discard treatments that are clearly not optimal.
  - Marginal value of observations for a given treatment is non-monotonic.

Introduction

Literature

- Multi-armed bandits – related but different:
  - Goal is to maximize outcomes of experimental units (rather than choose policy after experiment).
  - Exploration-exploitation trade-off (we focus on "exploration").
  - Units come in sequentially (rather than in waves).
- Good reviews:
  - Gittins index (optimal solution to some bandit problems): Weber et al. (1992)
  - Adaptive designs in clinical trials: Berry (2006).
  - Regret bounds for bandit problems: Bubeck and Cesa-Bianchi (2012).
  - Reinforcement learning: Ghavamzadeh et al. (2015).
  - Thompson sampling: Russo et al. (2018).
- Empirical examples for our simulations: Bryan et al. (2014), Ashraf et al. (2010), Cohen et al. (2015)

Setup

- Waves $t = 1, \ldots, T$, sample sizes N_t.
- Treatment $D \in \{1, \ldots, k\}$, outcomes $Y \in \{0, 1\}$.
- Potential outcomes Y^d.
- Repeated cross-sections:
  $(Y_{it}^0, \ldots, Y_{it}^k)$ are i.i.d. across both i and t.
- Average potential outcome:

$$\theta^d = E[Y_{it}^d].$$

- Key choice variable:
  Number of units n_t^d assigned to $D = d$ in wave t.
- Outcomes:
  Number of units s_t^d having a "success" (outcome $Y = 1$).

Setup

Treatment assignment, outcomes, state space

- Treatment assignment in wave t: $\boldsymbol{n}_t = (n_t^1, \ldots, n_t^k)$.
- Outcomes of wave t: $\boldsymbol{s}_t = (s_t^1, \ldots, s_t^k)$.
- Cumulative versions:

$$M_t = \sum_{t' \leq t} N_{t'}, \qquad \boldsymbol{m}_t = \sum_{t' \leq t} \boldsymbol{n}_t, \qquad \boldsymbol{r}_t = \sum_{t' \leq t} \boldsymbol{s}_t.$$

- Relevant information for the experimenter in period $t+1$ is summarized by $\boldsymbol{m}_t$ and $\boldsymbol{r}_t$.
- Total trials for each treatment, total successes.

Setup

Design objective

- Policy objective $SW(d)$:
  Average outcome Y, net of the cost of treatment.
- Choose treatment d after the experiment is completed.
- Posterior expected social welfare:

$$SW(d) = E[\theta^d | \boldsymbol{m}_T, \boldsymbol{r}_T] - c^d,$$

  where c^d is the unit cost of implementing policy d.

Setup

Bayesian prior and posterior

- By definition, $Y^d|\theta \sim Ber(\theta^d)$.
- Prior: $\theta^d \sim Beta(\alpha_0^d, \beta_0^d)$, independent across d.
- Posterior after period t:

$$\begin{aligned}\theta^d|\boldsymbol{m}_t, \boldsymbol{r}_t &\sim Beta(\alpha_t^d, \beta_t^d)\\ \alpha_t^d &= \alpha_0^d + r_t^d\\ \beta_t^d &= \beta_0^d + m_t^d - r_t^d.\end{aligned}$$

- In particular,

$$SW(d) = \frac{\alpha_0^d + r_T^d}{\alpha_0^d + \beta_0^d + m_T^d} - c^d.$$

Optimal treatment assignment

Optimal assignment: Dynamic optimization problem

- Dynamic stochastic optimization problem:
  - States $(\boldsymbol{m}_t, \boldsymbol{r}_t)$,
  - actions $\boldsymbol{n}_t$.
- Solve for the optimal experimental design using backward induction.
- Denote by V_t the value function after completion of wave t.
- Starting at the end, we have

$$V_T(\boldsymbol{m}_T, \boldsymbol{r}_T) = \max_d \left(\frac{\alpha_0^d + r_T^d}{\alpha_0^d + \beta_0^d + m_T^d} - c^d \right).$$

- Finite state and action space.
  $\Rightarrow$ Can, in principle, solve directly for optimal rule.
- But: Computation time quickly explodes.

Optimal treatment assignment

Simple examples

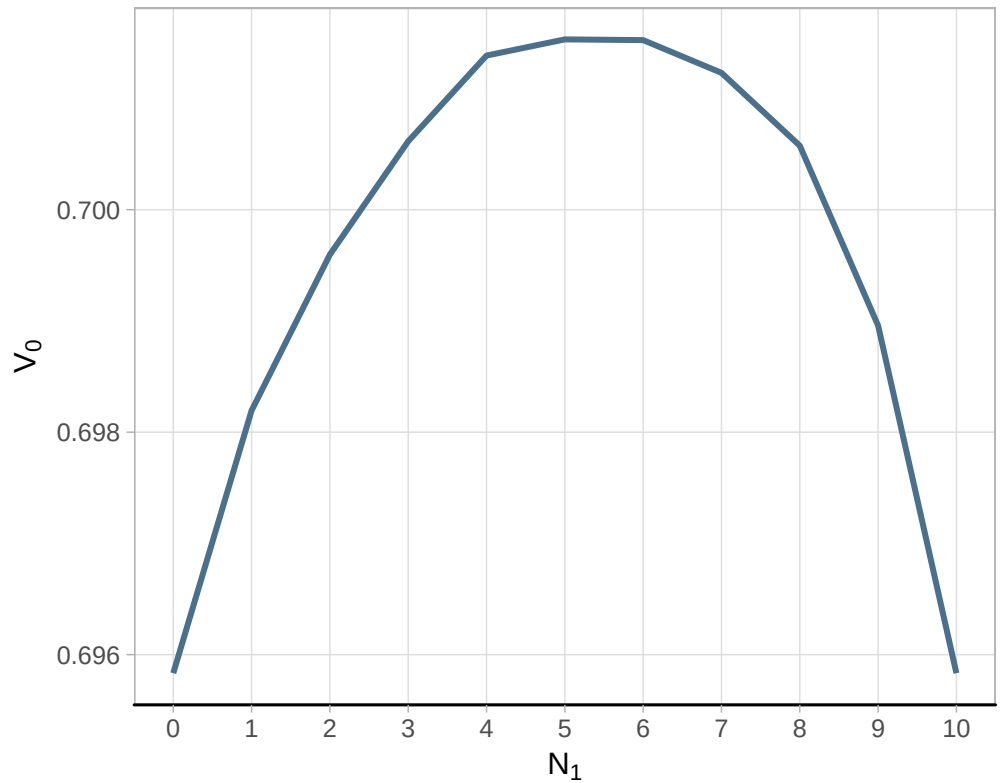
- Consider a small experiment
  with 2 waves, 3 treatment values (minimal interesting case).
- The following slides plot expected welfare
  as a function of:
  1. **Division of sample** size between waves, $N_1 + N_2 = 10$.
     $N_1 = 6$ is optimal.
  2. **Treatment assignment** in wave 2, given wave 1 outcomes.
     $N_1 = 6$ units in wave 1, $N_2 = 4$ units in wave 2.
- Keep in mind:

$$\alpha_1 = (1, 1, 1) + \boldsymbol{s}_1$$
$$\beta_1 = (1, 1, 1) + \boldsymbol{n}_1 - \boldsymbol{s}_1$$

Optimal treatment assignment

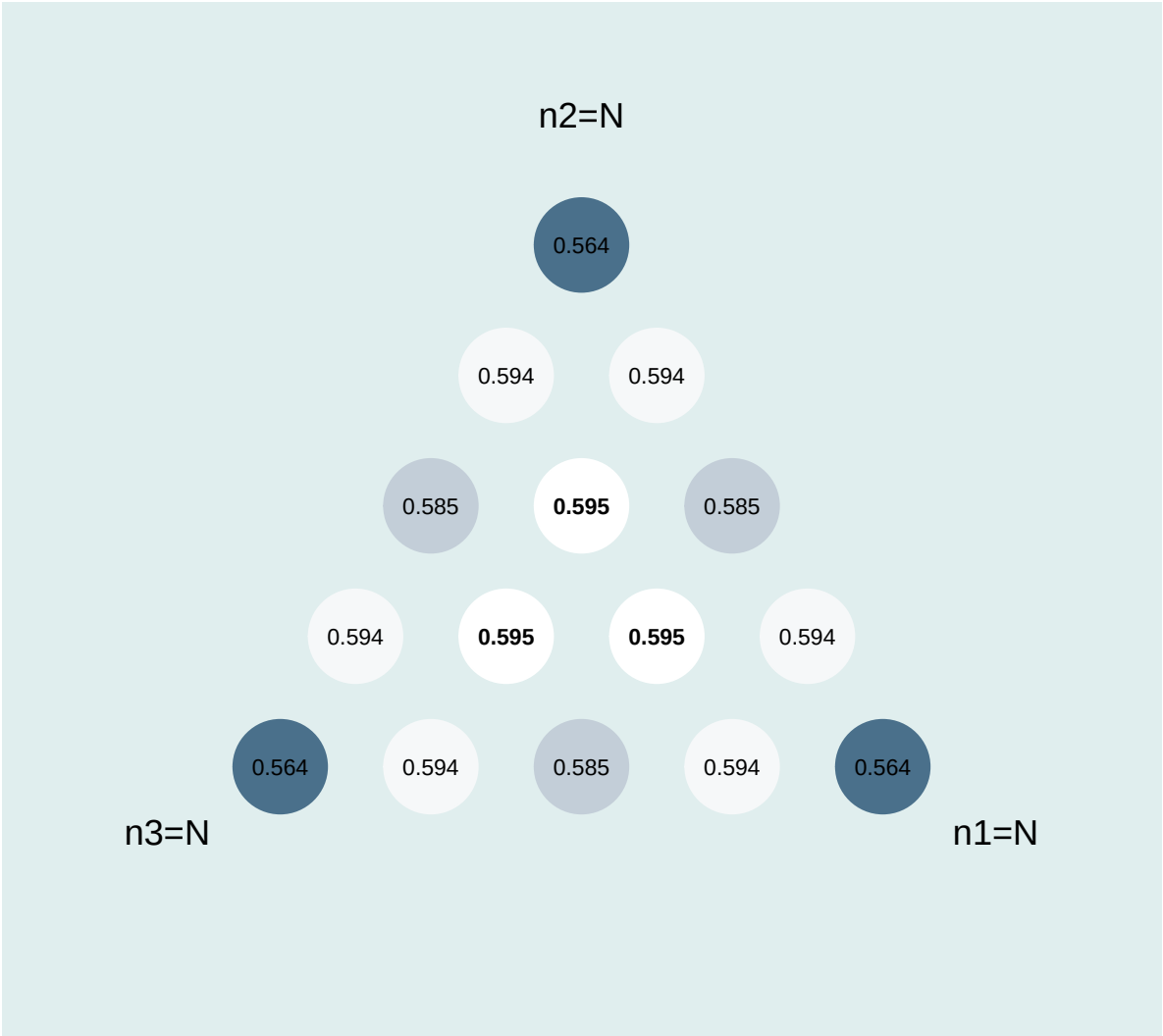
Dividing sample size between waves

- $N_1 + N_2 = 10$.
- Expected welfare as a function of N_1.
- Boundary points $\approx$ 1-wave experiment.
- $N_1 = 6$ (or 5) is optimal.



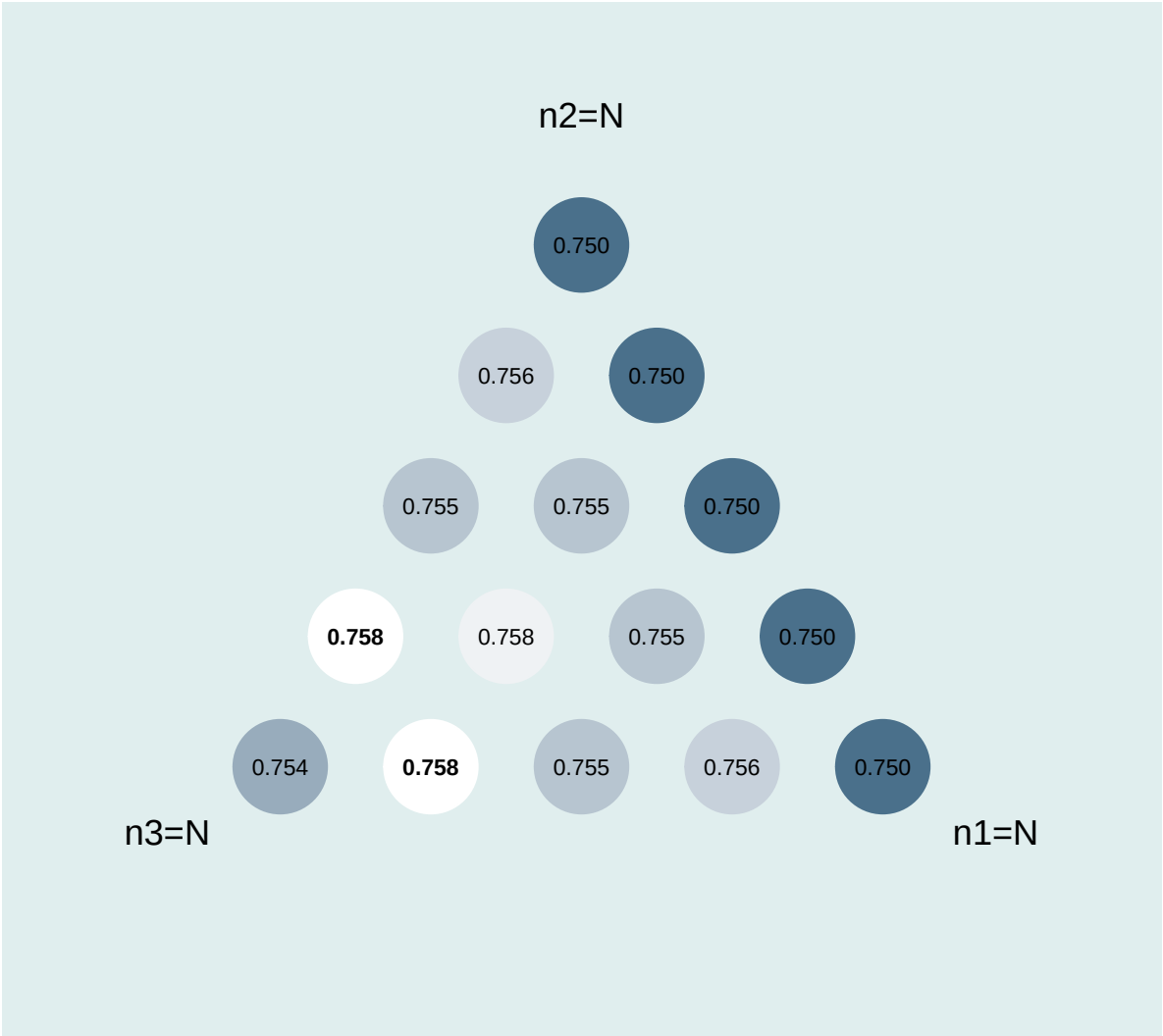
Optimal treatment assignment

$\alpha = (2, 2, 2)$, $\beta = (2, 2, 2)$



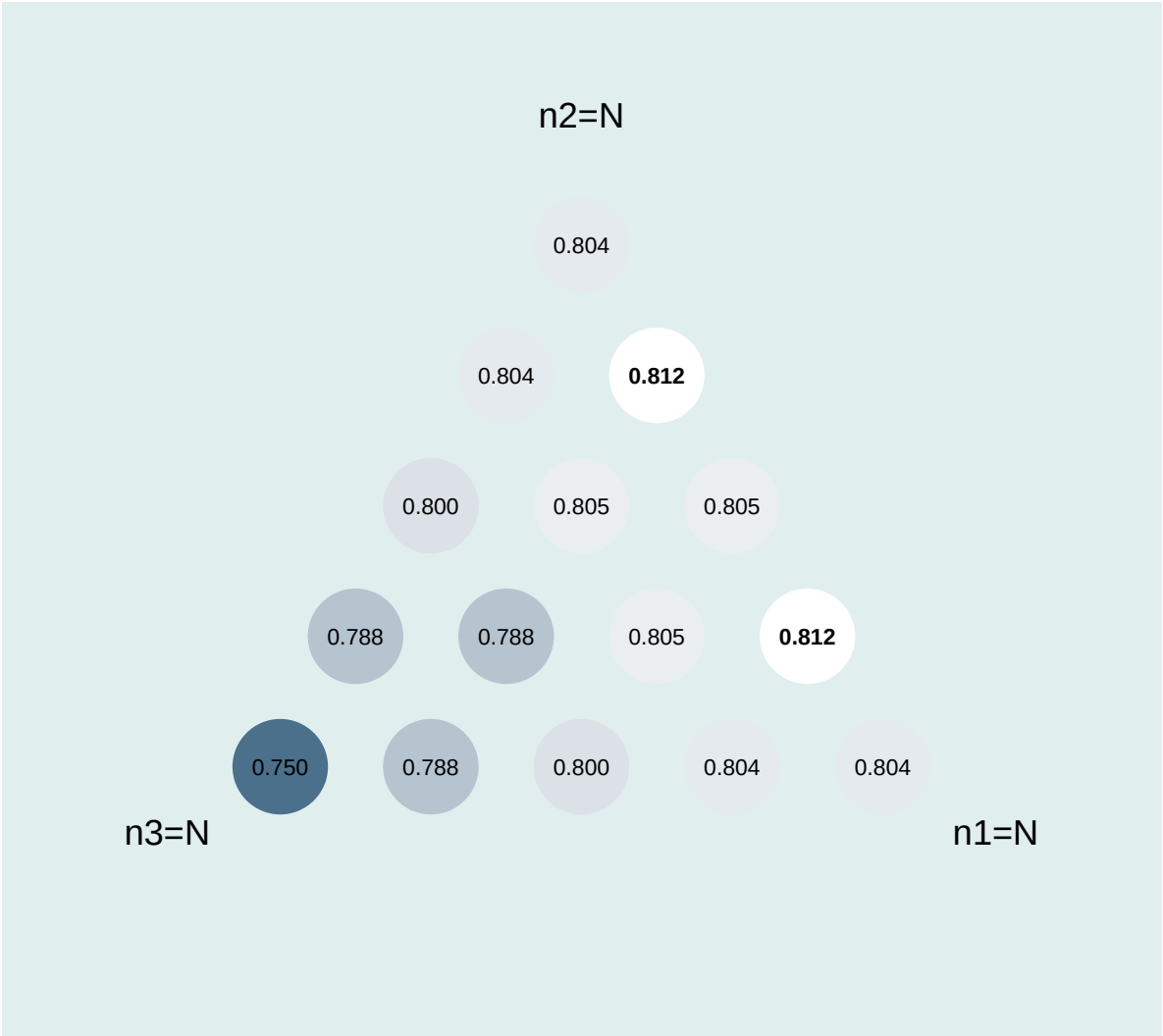
Optimal treatment assignment

$\alpha = (2, 2, 3)$, $\beta = (2, 2, 1)$



Optimal treatment assignment

$\alpha = (3, 3, 1)$, $\beta = (1, 1, 3)$



Modified Thompson sampling

A simpler alternative

- Old proposal by Thompson (1933) for clinical trials; popular in online experimentation.
- Assign each treatment with probability equal to the posterior probability that it is optimal.
- Easily implemented: Sample draws $\widehat{\boldsymbol{\theta}}_{it}$ from the posterior, assign

$$D_{it} = \underset{d}{\text{argmax}}\ \hat{\theta}_{it}^{d}.$$

- We propose two **modifications**:
  1. Don't assign the same treatment twice in a row.
  2. Re-run the algorithm several times, and use average n_t^d for each treatment d.

Modified Thompson sampling

Justifications

1. Mimics the qualitative behavior of optimal assignment in examples.
2. Thompson sampling has strong theoretical justifications (regret bounds) in multi armed bandit setting.
3. Modifications motivated by differences in setting:
   a) No exploitation motive.
   b) Waves rather than sequential arrival.
4. Performs well in calibrated simulations (coming up).
5. Is easy to compute.
6. Is easy to adapt to more general models.

Modified Thompson sampling

Extension: Covariates and treatment targeting

- Suppose now that
  1. We additionally observe a (discrete) covariate X.
  2. The policy to be chosen can **target treatment** by X.
- Implications for experimental design?
  1. Simple solution: Treat each covariate cell as its separate experiment; all the above applies.
  2. Better solution: Set up a hierarchical Bayes model, to optimally combine information across treatment cells.
- Example of a **hierarchical Bayes** model:

$$
\begin{aligned}
Y^d | X = x, \theta^{dx}, (\alpha_0^d, \beta_0^d) &\sim Ber(\theta^{dx}) \\
\theta^{dx} | (\alpha_0^d, \beta_0^d) &\sim Beta(\alpha_0^d, \beta_0^d) \\
(\alpha_0^d, \beta_0^d) &\sim \pi,
\end{aligned}
$$

Modified Thompson sampling

Calibrated simulations

- Simulate data calibrated to estimates of 3 published experiments.
- Set $\boldsymbol{\theta}$ equal to observed average outcomes for each stratum and treatment.
- Total sample size same as original.

Ashraf, N., Berry, J., and Shapiro, J. M. (2010). Can higher prices stimulate product use? Evidence from a field experiment in Zambia.
American Economic Review, 100(5):2383–2413

Bryan, G., Chowdhury, S., and Mobarak, A. M. (2014). Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh.
Econometrica, 82(5):1671–1748

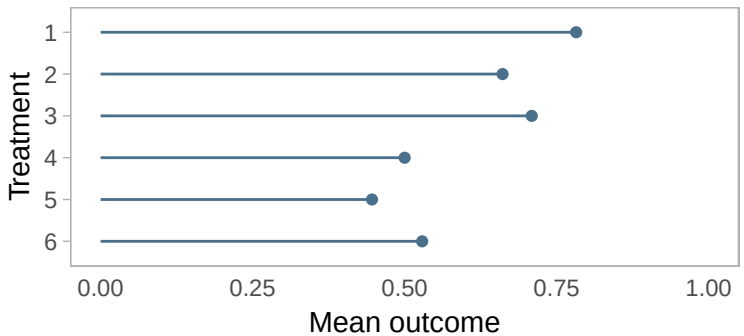
Cohen, J., Dupas, P., and Schaner, S. (2015). Price subsidies, diagnostic tests, and targeting of malaria treatment: evidence from a randomized controlled trial.
American Economic Review, 105(2):609–45

Modified Thompson sampling

Calibrated simulations – parameter values

6 treatments,
evenly spaced.

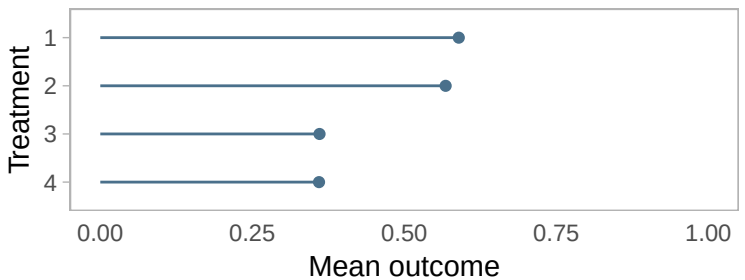
Ashraf, Berry, and Shapiro (2010)



Outcome: Whether the household purchased water disinfectant
Treatments: Subsidy levels from high to low.

2 close good treatments,
2 worse treatments.

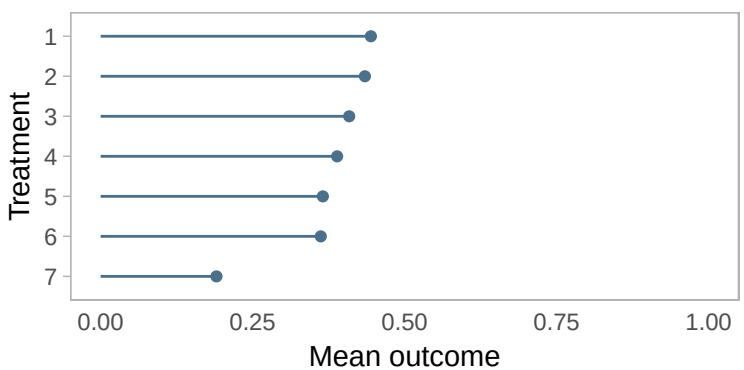
Bryan, Chowdhury, and Mobarak (2014)



Outcome: Whether at least one household member migrated
Treatments: Cash, credit, information, control group.

7 treatments,
closer than for first example.

Cohen, Dupas, and Schaner (2014)



Outcome: Bought ACT
Treatments: 3 subsidy levels with or without RDT, and control.

Modified Thompson sampling

Calibrated simulations - coming up

- Compare 4 **assignment methods**:
  1. Non-adaptive:
     Assign a share of $1/k$ of units to each treatment.
  2. Best half:
     Assign a share of $2/k$ of units to each of the $k/2$ treatments with highest posterior mean of θ^d.
  3. Thompson
  4. Modified Thompson
- Report 2 **statistics**:
  1. Regret:
     Average difference, across simulations, between $\max_d \theta^d$ and θ^d for the d chosen after the experiment.
  2. Share optimal:
     Share of simulations for which the optimal d is chosen after the experiment.

Modified Thompson sampling

Calibrated simulations, 2 waves

Table: 10000 replications, 2 waves.

Statistic	Ashraf	Bryan	Cohen
Regret, non-adaptive	0.005	0.005	0.009
Regret, best half	0.003	0.004	0.007
Regret, Thompson	0.003	0.005	0.007
Regret, modified Thompson	0.001	0.004	0.007
Share optimal, non-adaptive	0.929	0.748	0.525
Share optimal, best half	0.965	0.802	0.560
Share optimal, Thompson	0.963	0.776	0.548
Share optimal, modified Thompson	0.981	0.800	0.571
Units per wave	502	935	1080
Number of treatments	6	4	7

Modified Thompson sampling

Calibrated simulations, 4 waves

Table: 10000 replications, 4 waves.

Statistic	Ashraf	Bryan	Cohen
Regret, non-adaptive	0.005	0.005	0.009
Regret, best half	0.002	0.004	0.007
Regret, Thompson	0.002	0.005	0.007
Regret, modified Thompson	0.001	0.004	0.007
Share optimal, non-adaptive	0.929	0.767	0.525
Share optimal, best half	0.977	0.794	0.555
Share optimal, Thompson	0.977	0.787	0.578
Share optimal, modified Thompson	0.985	0.810	0.563
Units per wave	251	467	540
Number of treatments	6	4	7

Modified Thompson sampling

Calibrated simulations, 10 waves

Table: 10000 replications, 10 waves.

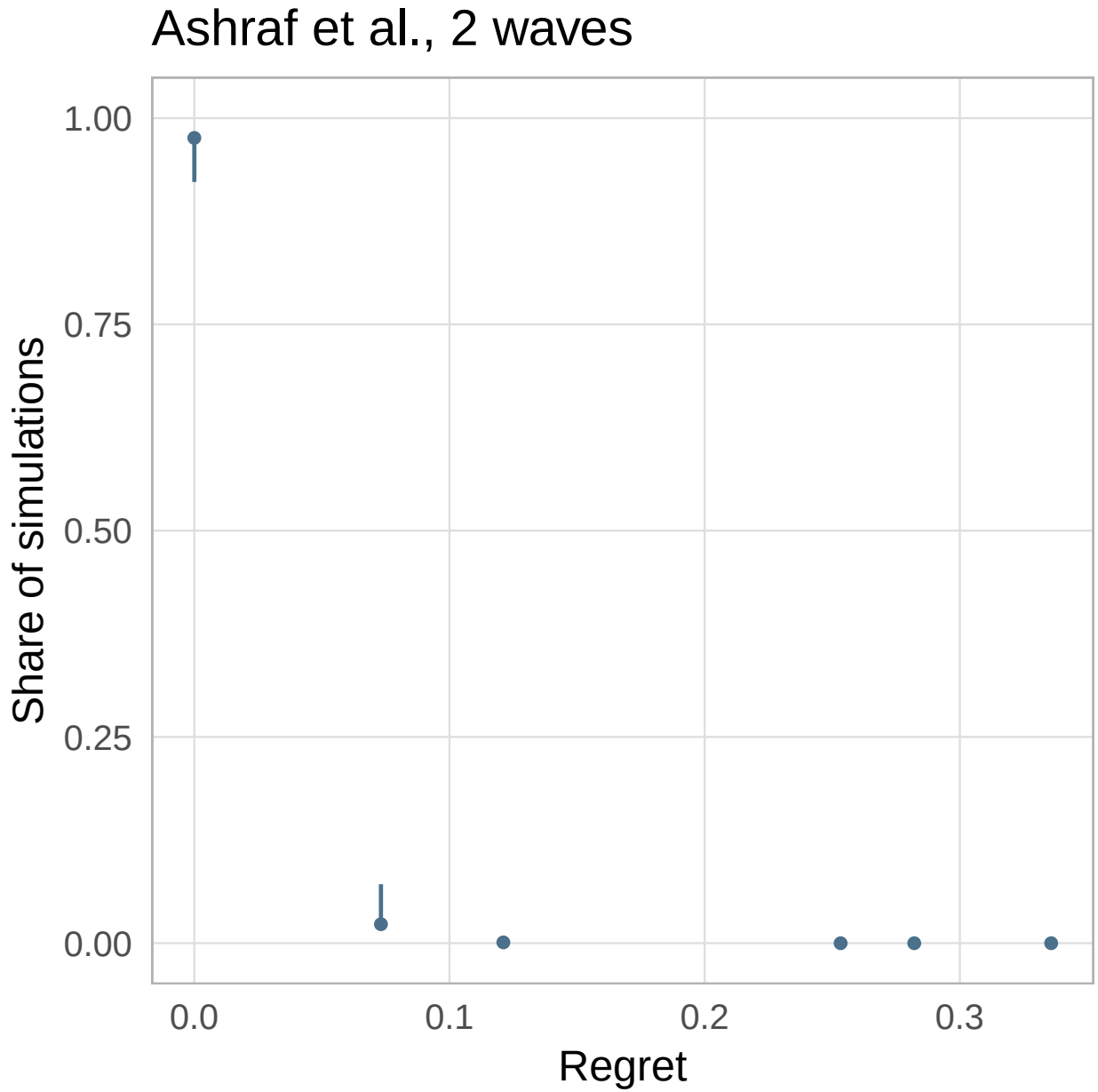
Statistic	Ashraf	Bryan	Cohen
regret, non-adaptive	0.005	0.005	0.009
regret, best half	0.002	0.004	0.007
regret, Thompson	0.001	0.004	0.006
regret, modified Thompson	0.001	0.004	0.006
share optimal, non-adaptive	0.939	0.749	0.530
share optimal, best half	0.977	0.820	0.560
share optimal, Thompson	0.981	0.811	0.601
share optimal, modified Thompson	0.988	0.819	0.596
units per wave	100	187	216
number of treatments	6	4	7

Modified Thompson sampling

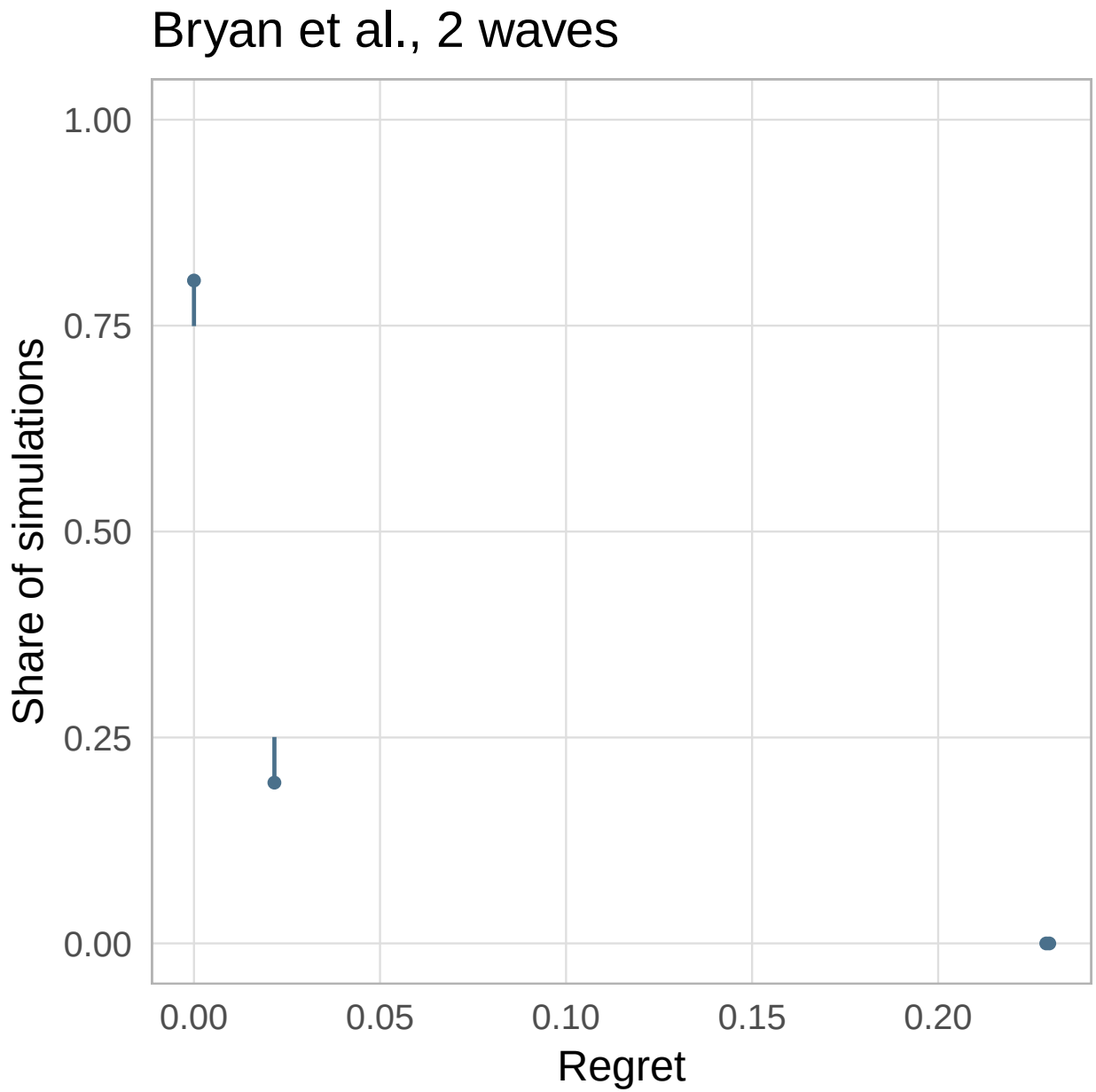
Calibrated simulations

- Next: visual representation of simulation results.
- Axes:
  - Horizontal: Regret of chosen policy after experiment.
  - Vertical: Share of simulations for which that policy was chosen.
- Comparing:
  - Modified Thompson sampling: Dot.
  - Non-adaptive design: other end of line.
- E.g.:
  - If dot is on top end of line for regret=0, then
  - the optimal treatment was chosen more often
    under modified Thompson sampling
    than under non-adaptive design.

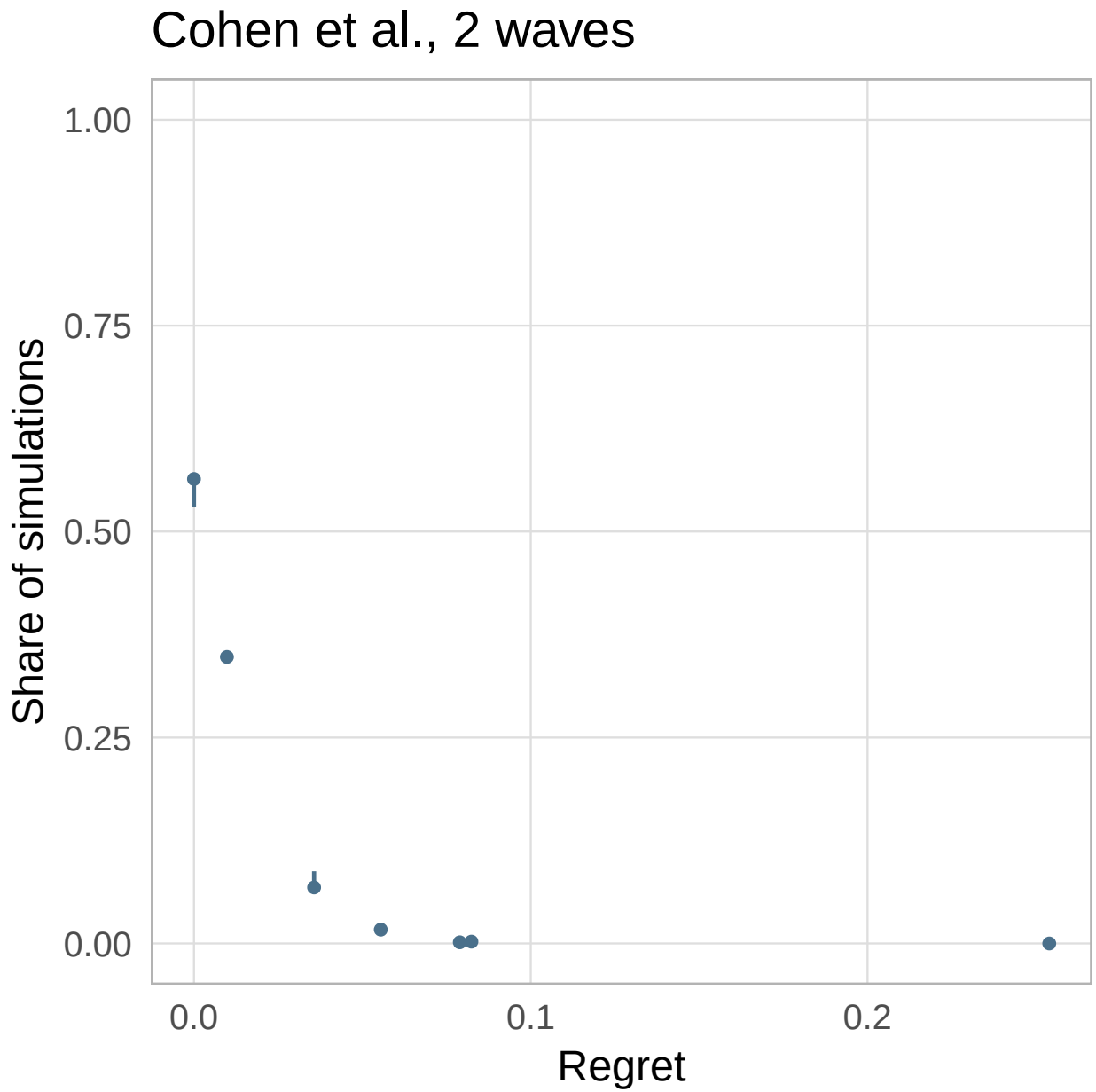
Modified Thompson sampling



Modified Thompson sampling



Modified Thompson sampling



Inference

- For inference, have to be careful with adaptive designs.
  1. **Standard inference** won't work:
     Sample means are biased, t-tests don't control size.
  2. But: **Bayesian** inference can ignore adaptiveness!
  3. **Randomization tests** can be modified to work.
- Example to get intuition for bias:
  - Flip a fair coin.
  - If head, flip again, else stop.
  - Probability dist: 50% tail-stop, 25% head-tail, 25% head-head.
  - Expected share of heads?

$$.5 \cdot 0 + .25 \cdot .5 + .25 \cdot 1 = .375 \neq .5.$$

- Randomization inference:
  - Strong null hypothesis: $Y_i^1 = \ldots = Y_i^k$.
  - Under null, easy to re-simulate treatment assignment.
  - Re-calculate test statistic each time.
  - Take $1 - \alpha$ quantile across simulations as critical value.

Conclusion

- The goal of many field experiments is to inform policy choice.
- Experimental designs that are good for treatment effect estimation, or power, are not optimal for policy choice.
- If the experiment can be implemented in multiple waves, adaptive designs for policy choice
  1. significantly increase welfare,
  2. by focusing attention on the best performing policy options in later waves.
- Implementation of our proposed procedure is easy, and easily adapted to new settings.

A web-app for implementing the proposed designs is available at

`https://maxkasy.shinyapps.io/ThompsonHierarchical/`

Conclusion

Questions for you

1. We are looking for field settings to implement our proposal. Suggestions?
2. Which directions should we push this?
   a) Theoretical characterizations of Thompson sampling?
   b) More simulations?
   c) More on inference?
   d) Hands-on cookbook?
   e) ...?

Thank you!