What do we want? And when do we want it? Alternative objectives and their implications for experimental design.

Maximilian Kasy

March 2020

Experimental design as a decision problem

How to assign treatments,
given the available information and objective?

Key ingredients when defining a decision problem:

1. **Objective function**:
   What is the ultimate goal? What will the experimental data be used for?
2. **Action space**:
   What information can experimental treatment assignments depend on?
3. **How to solve** the problem:
   Full optimization? Heuristic solution?
4. **How to evaluate** a solution:
   Risk function, Bayes risk, or worst case risk?

Experimental design as a decision problem

How to assign treatments,
given the available information and objective?

Key ingredients when defining a decision problem:

1. **Objective function**:
   What is the ultimate goal? What will the experimental data be used for?
2. **Action space**:
   What information can experimental treatment assignments depend on?
3. **How to solve** the problem:
   Full optimization? Heuristic solution?
4. **How to evaluate** a solution:
   Risk function, Bayes risk, or worst case risk?

Experimental design as a decision problem

How to assign treatments,
given the available information and objective?

Key ingredients when defining a decision problem:

1. **Objective function**:
   What is the ultimate goal? What will the experimental data be used for?
2. **Action space**:
   What information can experimental treatment assignments depend on?
3. **How to solve** the problem:
   Full optimization? Heuristic solution?
4. **How to evaluate** a solution:
   Risk function, Bayes risk, or worst case risk?

Experimental design as a decision problem

How to assign treatments,
given the available information and objective?

Key ingredients when defining a decision problem:

1. **Objective function**:
   What is the ultimate goal? What will the experimental data be used for?
2. **Action space**:
   What information can experimental treatment assignments depend on?
3. **How to solve** the problem:
   Full optimization? Heuristic solution?
4. **How to evaluate** a solution:
   Risk function, Bayes risk, or worst case risk?

Four possible types of objective functions for experiments

1. **Squared error** for estimates.
   - For instance for the average treatment effect.
   - Possibly weighted squared error of multiple estimates.
2. **In-sample average** outcomes.
   - Possibly transformed (inequality aversion),
   - costs taken into account, discounted.
3. **Policy choice** to maximize average **observed outcomes**.
   - Choose a policy after the experiment.
   - Evaluate the experiment based on the implied policy choice.
4. **Policy choice** to maximize **utilitarian welfare**.
   - Similar, but welfare is not directly observed.
   - Instead, maximize a weighted average (across people) of equivalent variation.

This talk:

- Review of several of my papers, considering each of these in turn.

Space of possible experimental designs

What information can treatment assignment condition on?

1. **Covariates**?
   $\Rightarrow$ Stratified and targeted treatment assignment.
2. **Earlier outcomes** for other units, in sequential or batched settings?
   $\Rightarrow$ Adaptive treatment assignment.

This talk:

- First conditioning on covariates,
  then settings without conditioning (for exposition only).
- First non-adaptive,
  then adaptive experiments.

Two approaches to optimization

1. **Fully optimal** designs.
   - Conceptually straightforward (dynamic stochastic optimization), but numerically challenging.
   - Preferred in the economic theory literature, which has focused on tractable (but not necessarily practically relevant) settings.
   - Do not require randomization.
2. **Approximately optimal** or rate optimal designs.
   - Heuristic algorithms.
   - Prove (rate)-optimality ex post.
   - Preferred in the machine learning literature. This is the approach that has revived the bandit literature and made it practically relevant.
   - Might involve randomization.

This talk:

- Approximately optimal algorithms.
- Bayesian algorithms, but we characterize the *risk function*, i.e., behavior conditional on the true parameter.

This talk: Several papers considering different objectives...

- **Minimizing squared error**:
  Kasy, M. (2016).
  Why experimenters might not always want to randomize, and what they could do instead.
  Political Analysis, 24(3):324–338.
- **Maximizing in-sample outcomes**:
  Caria, S., Gordon, G., Kasy, M., Osman, S., Quinn, S., and Teytelboym, A. (2020).
  Job search assistance for refugees in Jordan: An adaptive field experiment.
  Work in progress.
- **Optimizing policy choice – average outcomes**:
  Kasy, M. and Sautmann, A. (2020).
  Adaptive treatment assignment in experiments for policy choice.
  Conditionally accepted at Econometrica

... and outlook

- **Optimizing policy choice – utilitarian welfare**:
  Kasy, M. (2020). Adaptive experiments for optimal taxation.
  building on

  Kasy, M. (2019).
  Optimal taxation and insurance using machine learning – sufficient statistics and beyond. *Journal of Public Economics*.
- **Combinatorial allocation** (e.g. matching):
  Kasy, M. and Teytelboym, A. (2020a).
  Adaptive combinatorial allocation under constraints.
  Work in progress.
- **Testing in a pandemic**:
  Kasy, M. and Teytelboym, A. (2020b).
  Adaptive targeted disease testing.
  Forthcoming, Oxford Review of Economic Policy.

Literature

- Statistical decision theory:
  Berger (1985),
  Robert (2007).
- Non-parametric Bayesian methods:
  Ghosh and Ramamoorthi (2003),
  Williams and Rasmussen (2006),
  Ghosal and Van der Vaart (2017).
- Stratification and re-randomization:
  Morgan and Rubin (2012),
  Athey and Imbens (2017).
- Adaptive designs in clinical trials:
  Berry (2006),
  FDA et al. (2018).
- Bandit problems:
  Weber et al. (1992),
  Bubeck and Cesa-Bianchi (2012),
  Russo et al. (2018).
- Regret bounds:
  Agrawal and Goyal (2012),
  Russo and Van Roy (2016).
- Best arm identification:
  Glynn and Juneja (2004),
  Bubeck et al. (2011),
  Russo (2016).
- Bayesian optimization:
  Powell and Ryzhov (2012),
  Frazier (2018).
- Reinforcement learning:
  Ghavamzadeh et al. (2015),
  Sutton and Barto (2018).
- Optimal taxation:
  Mirrlees (1971),
  Saez (2001),
  Chetty (2009),
  Saez and Stantcheva (2016).

Minimizing squared error

Maximizing in-sample outcomes

Optimizing policy choice: Average outcomes

Outlook

- Utilitarian welfare
- Combinatorial allocation
- Testing in a pandemic

Conclusion and summary

No randomization in general decision problems

Theorem (Optimality of deterministic decisions)

Consider a general decision problem.
Let $R^*(\cdot)$ equal either Bayes risk or worst case risk. Then:

1. *The optimal risk $R^*(\delta^*)$, when considering only deterministic procedures is no larger than the optimal risk when allowing for randomized procedures.*
2. *If the optimal deterministic procedure is unique, then it has strictly lower risk than any non-trivial randomized procedure.*

Sketch of proof (Kasy, 2016):

- The risk function of a randomized procedure is a weighted average of the risk functions of deterministic procedures.
- The same is true for Bayes risk and minimax risk.
- The lowest risk is (weakly) smaller than the weighted average.

Minimizing squared error: Setup

1. **Sampling:** Random sample of n units.
   Baseline survey $\Rightarrow$ vector of covariates X_i.
2. **Treatment assignment:** Binary treatment assigned by $D_i = d_i(\boldsymbol{X}, U)$.
   $\boldsymbol{X}$ matrix of covariates; U randomization device .
3. **Realization of outcomes:** $Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$
4. **Estimation:** Estimator $\widehat{\beta}$ of the (conditional) average treatment effect,
   $\beta = \frac{1}{n} \sum_i E[Y_i^1 - Y_i^0 | X_i, \theta]$

Prior:

- Let $f(x, d) = E[Y_i^d | X_i = x]$.
- Let $C((x, d), (x', d'))$ be the prior covariance of $f(x, d)$ and $f(x', d')$.
- E.g. Gaussian process prior $f \sim GP(0, C(\cdot, \cdot))$.

Expected squared error

- Notation:
    - C: $n \times n$ prior covariance matrix of the $f(X_i, D_i)$.
    - $\bar{C}$: n vector of prior covariances of $f(X_i, D_i)$ with the CATE β.
    - $\widehat{\beta}$: The posterior best linear predictor of β.
- Kasy (2016):
  The Bayes **risk** (expected squared error) of a treatment assignment equals

$$\mathrm{Var}(\beta|\boldsymbol{X}) - \overline{C}' \cdot (C + \sigma^2 I)^{-1} \cdot \overline{C},$$

  where the prior variance $\mathrm{Var}(\beta|\boldsymbol{X})$ does not depend on the assignment, but $\overline{C}$ and C do.

Optimal design

- The **optimal design** minimizes the Bayes risk (expected squared error).
- For continuous covariates, the optimum is generically unique, and a non-random assignment is optimal.
- Expected squared error is a measure of **balance** across treatment arms.
- Simple approximate optimization algorithm: Re-randomization.

Two Caveats:

- Randomization inference requires randomization – outside of decision theory.
- If minimizing worst case risk given procedure, but not given randomization, mixed strategies can be optimal (Banerjee et al., 2017).

Maximizing in-sample outcomes

- Minimizing **squared error** is appropriate when you want to get precise estimates of policy effects.
- But in many settings we want to also **help participants** as much as possible.
- As argued by Kant (1791):

  Act in such a way that you treat humanity, whether in your own person or in the person of any other, never merely as a means to an end, but always at the same time as an end.

- If we care about both participant welfare and estimator precision, we might try to **trade both off**.
- This is done by the γ-Thompson algorithm that I will introduce shortly.

Adaptive targeted assignment: Setup

- Waves $t = 1, \ldots, T$, sample sizes N_t.
- Treatment $D \in \{1, \ldots, k\}$, outcomes $Y \in [0,1]$, covariate $X \in \{1, \ldots, n_x\}$.
- Potential outcomes Y^d.
- Repeated cross-sections: $(Y_{it}^1, \ldots, Y_{it}^k, X_{it})$ are i.i.d. across both i and t.
- Average potential outcomes:
$$\theta^{dx} = E[Y_{it}^d | X_{it} = x].$$
- **Regret**: Difference in average outcomes from decision d versus the optimal decision,
$$\Delta^{dx} = \max_{d'} \theta^{d'x} - \theta^{dx}.$$
- Average in-sample regret:
$$\frac{1}{\sum_t N_t} \sum_{i,t} \Delta^{D_{it} X_{it}}.$$

Thompson sampling and γ-Thompson sampling

- **Thompson sampling**
    - Old proposal by Thompson (1933).
    - Popular in online experimentation.
- Assign each treatment with probability equal to the posterior probability that it is optimal, given $X = x$ and given the information available at time t.

$$p_t^{dx} = P_t\left(d = \underset{d'}{\operatorname{argmax}}\ \theta^{d'x}\right).$$

- γ-**Thompson sampling**: Assign each treatment with probability equal to

$$(1-\gamma)\cdot p_t^{dx} + \gamma/k.$$

Compromise between full randomization and Thompson sampling.

My development economics co-authors want to both publish estimates and help!

Limiting behavior

Theorem (Caria et al. 2020)

Given $\boldsymbol{\theta}$, as $t \to \infty$:

1. *The* **cumulative share** q_t^{dx} *allocated to treatment d in stratum x converges in probability to $\bar{q}^{dx} = (1-\gamma) + \gamma/k$ for $d = d^{*x}$, and to $\bar{q}^{dx} = \gamma/k$ for all other d.*
2. *Average* **in-sample regret** *converges in probability to*
$$\gamma \cdot \left(\frac{1}{k} \sum_{x,d} \Delta^{dx} \cdot p^x \right).$$
3. *The normalized* **average outcome** *for treatment d in stratum x, $\sqrt{M_t}\left(\bar{Y}_t^{dx} - \theta_0^{dx}\right)$, converges in distribution to*
$$N\left(0, \frac{\theta_0^{dx}(1-\theta_0^{dx})}{\bar{q}^{dx} \cdot p^x}\right).$$

Limiting behavior

Theorem (Caria et al. 2020)

Given $\boldsymbol{\theta}$, as $t \to \infty$:

1. *The* **cumulative share** q_t^{dx} *allocated to treatment d in stratum x converges in probability to $\bar{q}^{dx} = (1-\gamma) + \gamma/k$ for $d = d^{*x}$, and to $\bar{q}^{dx} = \gamma/k$ for all other d.*
2. *Average* **in-sample regret** *converges in probability to*

$$\gamma \cdot \left(\frac{1}{k} \sum_{x,d} \Delta^{dx} \cdot p^x \right).$$

3. *The normalized* **average outcome** *for treatment d in stratum x, $\sqrt{M_t}\left(\bar{Y}_t^{dx} - \theta_0^{dx}\right)$, converges in distribution to*

$$N\left(0, \frac{\theta_0^{dx}(1-\theta_0^{dx})}{\bar{q}^{dx} \cdot p^x}\right).$$

Limiting behavior

Theorem (Caria et al. 2020)

Given $\boldsymbol{\theta}$, as $t \to \infty$:

1. *The* **cumulative share** q_t^{dx} *allocated to treatment d in stratum x converges in probability to $\bar{q}^{dx} = (1-\gamma) + \gamma/k$ for $d = d^{*x}$, and to $\bar{q}^{dx} = \gamma/k$ for all other d.*
2. *Average* **in-sample regret** *converges in probability to*

$$\gamma \cdot \left(\frac{1}{k} \sum_{x,d} \Delta^{dx} \cdot p^x \right).$$

3. *The normalized* **average outcome** *for treatment d in stratum x, $\sqrt{M_t}\left(\bar{Y}_t^{dx} - \theta_0^{dx}\right)$, converges in distribution to*

$$N\left(0, \frac{\theta_0^{dx}(1-\theta_0^{dx})}{\bar{q}^{dx} \cdot p^x}\right).$$

Interpretation

- **In-sample regret** is (approximately) proportional to the share γ of observations fully randomized.
- The **variance** of average potential outcome estimators is proportional
  - to $\frac{1}{\gamma/k}$ for sub-optimal d,
  - to $\frac{1}{(1-\gamma)+\gamma/k}$ for conditionally optimal d.
- The variance of **treatment effect** estimators, comparing the conditional optimum to alternatives, is therefore decreasing in γ.
- An **optimal** choice of γ could **trade off** regret and estimator variance.

In the application coming next, we chose $\gamma = .2$, somewhat arbitrarily.

Application: Job search assistance for refugees in Jordan

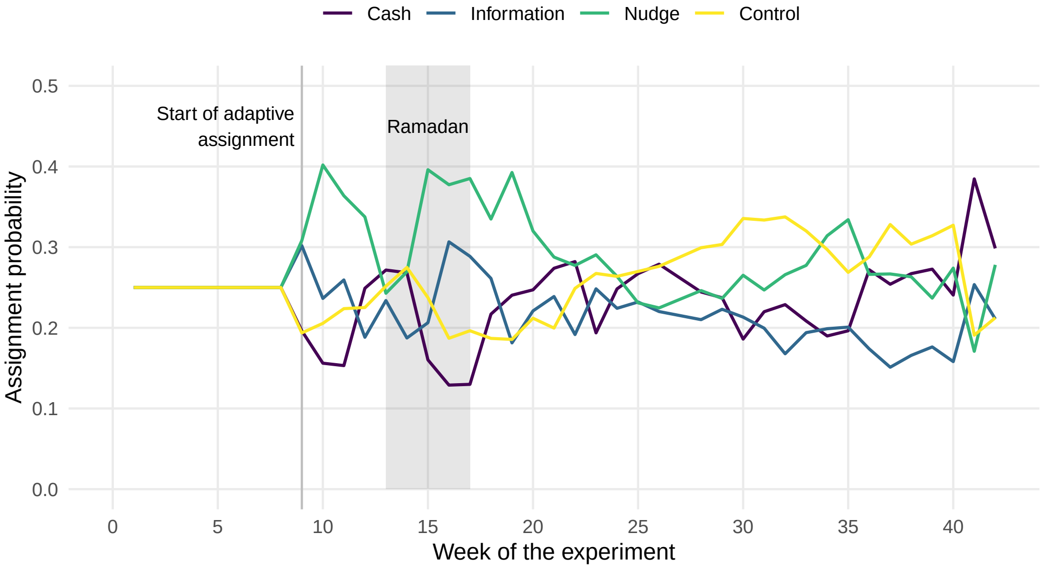
- Jordan 2019, International Rescue Committee.
  - Participants: Syrian refugees and Jordanians.
  - Main locations: Amman and Irbid.
  - Sample size: 3770.
- **Context**: Jordan compact.
  Gave refugees the right to work in low-skilled formal jobs.
- 4 **Treatments**:
  1. Cash: 65 JOD (91.5 USD).
  2. Information: On (i) how to interview for a formal job, and (ii) labor law and worker rights.
  3. Nudge: A job-search planning session and SMS reminders.
  4. Control group.
- **Conditioning variables** for treatment assignment: 16 strata, based on
  1. nationality (Jordanian or Syrian),
  2. gender,
  3. education (completed high school or more), and
  4. work experience (having experience in wage employment).

Locations

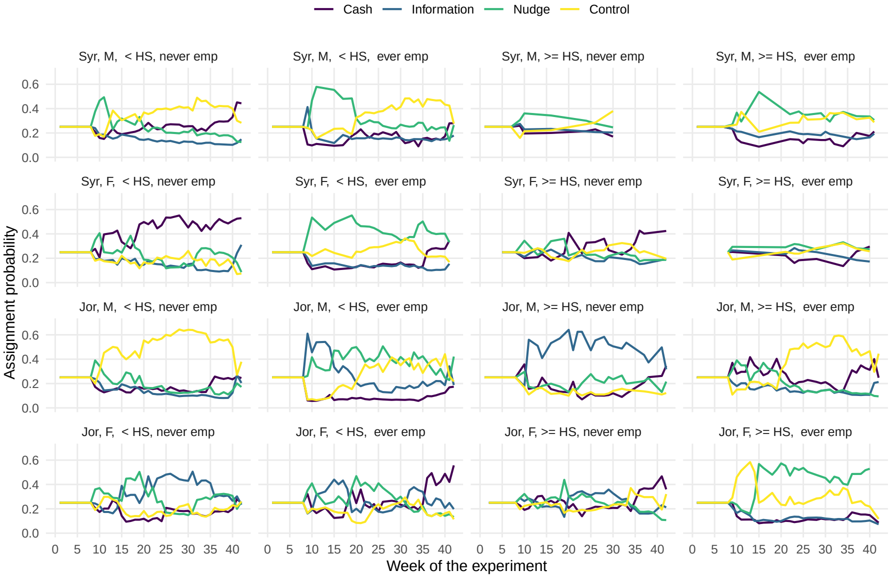
Irbid

Amman

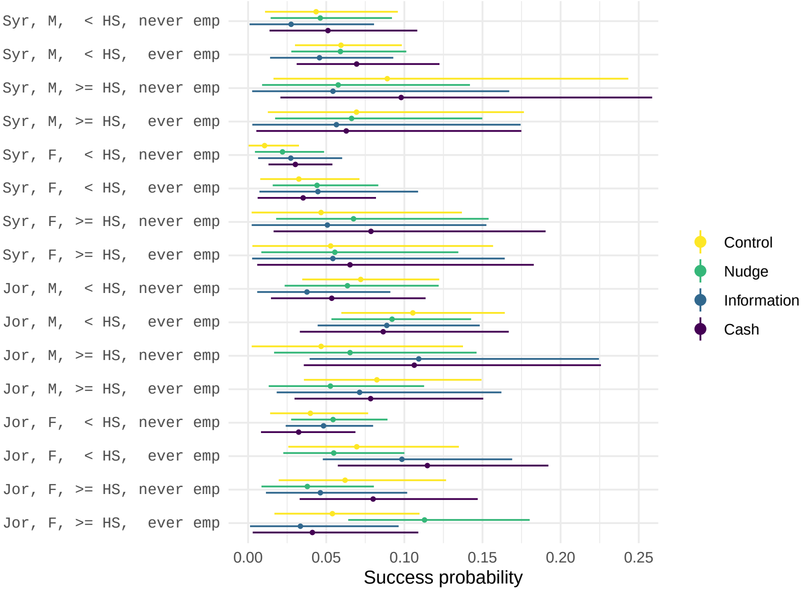
Assignment probabilities over time



Assignment probabilities over time, by stratum



Effect heterogeneity: Posterior means and 95% credible sets



Minimizing squared error

Maximizing in-sample outcomes

Optimizing policy choice: Average outcomes

Outlook

- Utilitarian welfare
- Combinatorial allocation
- Testing in a pandemic

Conclusion and summary

Optimizing policy choice: Average outcomes

- Setup: As before, but without covariates (just for presentation).
- Suppose you will **choose a policy** after the experiment, based on posterior beliefs,

$$d_T^* \in \underset{d}{\operatorname{argmax}}\ \hat{\theta}_T^d, \qquad \hat{\theta}_T^d = E_T[\theta^d].$$

- Evaluate experimental designs based on expected welfare (ex ante, given $\boldsymbol{\theta}$).
- Equivalently, **expected policy regret**

$$\mathrm{R(T)} = \sum_d \Delta^d \cdot P\left(d_T^* = d\right), \qquad \Delta^d = \max_{d'} \theta^{d'} - \theta^d.$$

- **Justification**:
  - Continuing experimentation is costly and requires oversight.
  - Political constraints might prevent indefinite experimentation.
  - Experimental samples are often small relative to the policy-population.

The infeasible rate-optimal allocation

- For good designs, R(T) converges to 0 at a fast rate.
- We can characterize the oracle-optimal shares $\bar{q}^d$ allocated to each treatment d, given $\boldsymbol{\theta}$, as follows:

1. The **rate** of convergence to 0 of **policy regret** $\mathrm{R(T)} = \sum_d \Delta^d \cdot P\left(d_T^* = d\right)$ is equal to the slowest rate of convergence of $P\left(d_T^* = d\right)$ across the sub-optimal d.
2. The **rate** of convergence of the **probability** $P\left(d_T^* = d\right)$ is increasing in the share $\bar{q}^d$ assigned to d, and is also increasing in the effect size Δ^d. It is equal to the rate of convergence of the posterior probability p_t^d.
3. The **optimal sample shares** $\bar{q}^d$ equalize the rate of convergence of $P\left(d_T^* = d\right)$ across sub-optimal d.

This is infeasible, since it requires knowledge of θ!

The infeasible rate-optimal allocation

- For good designs, R(T) converges to 0 at a fast rate.
- We can characterize the oracle-optimal shares $\bar{q}^d$ allocated to each treatment d, given $\boldsymbol{\theta}$, as follows:

1. The **rate** of convergence to 0 of **policy regret** $\mathrm{R(T)} = \sum_d \Delta^d \cdot P\left(d_T^* = d\right)$ is equal to the slowest rate of convergence of $P\left(d_T^* = d\right)$ across the sub-optimal d.
2. The **rate** of convergence of the **probability** $P\left(d_T^* = d\right)$ is increasing in the share $\bar{q}^d$ assigned to d, and is also increasing in the effect size Δ^d. It is equal to the rate of convergence of the posterior probability p_t^d.
3. The **optimal sample shares** $\bar{q}^d$ equalize the rate of convergence of $P\left(d_T^* = d\right)$ across sub-optimal d.

   This is infeasible, since it requires knowledge of θ!

The infeasible rate-optimal allocation

- For good designs, R(T) converges to 0 at a fast rate.
- We can characterize the oracle-optimal shares $\bar{q}^d$ allocated to each treatment d, given $\boldsymbol{\theta}$, as follows:

1. The **rate** of convergence to 0 of **policy regret** $\mathrm{R(T)} = \sum_d \Delta^d \cdot P\left(d_T^* = d\right)$ is equal to the slowest rate of convergence of $P\left(d_T^* = d\right)$ across the sub-optimal d.
2. The **rate** of convergence of the **probability** $P\left(d_T^* = d\right)$ is increasing in the share $\bar{q}^d$ assigned to d, and is also increasing in the effect size Δ^d. It is equal to the rate of convergence of the posterior probability p_t^d.
3. The **optimal sample shares** $\bar{q}^d$ equalize the rate of convergence of $P\left(d_T^* = d\right)$ across sub-optimal d.

   This is infeasible, since it requires knowledge of $\boldsymbol{\theta}$!

Exploration sampling

- How do we construct a feasible algorithm that behaves in the same way?
- Agrawal and Goyal (2012) proved that Thompson-sampling is rate-optimal for the multi-armed bandit problem. It is not for our policy choice problem!
- We propose the following modification.
- **Exploration sampling**:
  Assign shares q_t^d of each wave to treatment d, where

$$q_t^d = S_t \cdot p_t^d \cdot (1 - p_t^d),$$

$$p_t^d = P_t \left(d = \underset{d'}{\operatorname{argmax}} \ \theta^{d'} \right), \qquad S_t = \frac{1}{\sum_d p_t^d \cdot (1 - p_t^d)}.$$

- This modification
  1. yields rate-optimality (theorem coming up), and
  2. improves performance in our simulations.

Exploration sampling is rate optimal

Theorem (Kasy and Sautmann 2020)

Consider exploration sampling in a setting with fixed wave size ≥ 1.
Assume that $\max_d \theta^d < 1$ *and that the optimal policy* $\operatorname{argmax}_d \theta^d$ *is unique.*
As $T \to \infty$, the following holds:

1. *The share of observations assigned to the best treatment converges in probability to $1/2$.*
2. *The share of observations assigned to treatment d for all other d converges in probability to a non-random share $\bar{q}^d$.*
   $\bar{q}^d$ is such that $-\frac{1}{NT} \log p_t^d \to^p \Gamma^*$ for some $\Gamma^* > 0$ that is constant across $d \neq \operatorname{argmax}_d \theta^d$.
3. *Expected policy regret converges to 0 at the same rate Γ^*, that is, $-\frac{1}{NT} \log \mathrm{R}(\mathrm{T}) \to^p \Gamma^*$.*
   No other assignment shares $\bar{q}^d$ exist for which 1. holds and $\mathrm{R}(\mathrm{T})$ goes to 0 at a faster rate than Γ^*.

Exploration sampling is rate optimal

Theorem (Kasy and Sautmann 2020)

Consider exploration sampling in a setting with fixed wave size ≥ 1.
Assume that $\max_d \theta^d < 1$ *and that the optimal policy* $\operatorname{argmax}_d \theta^d$ *is unique.*
As $T \to \infty$, the following holds:

1. *The share of observations assigned to the best treatment converges in probability to $1/2$.*
2. *The share of observations assigned to treatment d for all other d converges in probability to a non-random share $\bar{q}^d$.*
   $\bar{q}^d$ *is such that* $-\frac{1}{NT} \log p_t^d \to^p \Gamma^*$
   for some $\Gamma^* > 0$ that is constant across $d \neq \operatorname{argmax}_d \theta^d$.
3. *Expected policy regret converges to 0 at the same rate Γ^*, that is,* $-\frac{1}{NT} \log R(T) \to^p \Gamma^*$.
   No other assignment shares $\bar{q}^d$ exist for which 1. holds and $R(T)$ *goes to 0 at a faster rate than Γ^*.*

Exploration sampling is rate optimal

Theorem (Kasy and Sautmann 2020)

Consider exploration sampling in a setting with fixed wave size ≥ 1.
Assume that $\max_d \theta^d < 1$ *and that the optimal policy* $\operatorname{argmax}_d \theta^d$ *is unique.*
As $T \to \infty$*, the following holds:*

1. *The share of observations assigned to the best treatment converges in probability to* $1/2$.
2. *The share of observations assigned to treatment* d *for all other* d *converges in probability to a non-random share* $\bar{q}^d$.
   $\bar{q}^d$ *is such that* $-\frac{1}{NT}\log p_t^d \to^p \Gamma^*$ *for some* $\Gamma^* > 0$ *that is constant across* $d \neq \operatorname{argmax}_d \theta^d$.
3. *Expected policy regret converges to* 0 *at the same rate* Γ^**, that is,* $-\frac{1}{NT}\log \mathrm{R(T)} \to^p \Gamma^*$.
   No other assignment shares $\bar{q}^d$ *exist for which 1. holds and* $\mathrm{R(T)}$ *goes to* 0 *at a faster rate than* Γ^*.

Sketch of proof

Our proof draws on several Lemmas of Glynn and Juneja (2004) and Russo (2016).

Proof steps:

1. Each treatment is assigned infinitely often.
   $\Rightarrow p_T^d$ goes to 1 for the optimal treatment and to 0 for all other treatments.
2. Claim 1 then follows from the definition of exploration sampling.
3. Claim 2: Suppose p_t^d goes to 0 at a faster rate for some d.
   Then exploration sampling stops assigning this d.
   This allows the other treatments to "catch up."
4. Claim 3: Balancing the rate of convergence implies efficiency.
   This follows from the rate-optimal allocation discussed before.

Sketch of proof

Our proof draws on several Lemmas of Glynn and Juneja (2004) and Russo (2016).

Proof steps:

1. Each treatment is assigned infinitely often.
   $\Rightarrow p_T^d$ goes to 1 for the optimal treatment and to 0 for all other treatments.
2. Claim 1 then follows from the definition of exploration sampling.
3. Claim 2: Suppose p_t^d goes to 0 at a faster rate for some d.
   Then exploration sampling stops assigning this d.
   This allows the other treatments to "catch up."
4. Claim 3: Balancing the rate of convergence implies efficiency.
   This follows from the rate-optimal allocation discussed before.

Sketch of proof

Our proof draws on several Lemmas of Glynn and Juneja (2004) and Russo (2016).

Proof steps:

1. Each treatment is assigned infinitely often.
   $\Rightarrow p_T^d$ goes to 1 for the optimal treatment and to 0 for all other treatments.
2. Claim 1 then follows from the definition of exploration sampling.
3. Claim 2: Suppose p_t^d goes to 0 at a faster rate for some d.
   Then exploration sampling stops assigning this d.
   This allows the other treatments to "catch up."
4. Claim 3: Balancing the rate of convergence implies efficiency.
   This follows from the rate-optimal allocation discussed before.

Sketch of proof

Our proof draws on several Lemmas of Glynn and Juneja (2004) and Russo (2016).

Proof steps:

1. Each treatment is assigned infinitely often.
   $\Rightarrow p_T^d$ goes to 1 for the optimal treatment and to 0 for all other treatments.
2. Claim 1 then follows from the definition of exploration sampling.
3. Claim 2: Suppose p_t^d goes to 0 at a faster rate for some d.
   Then exploration sampling stops assigning this d.
   This allows the other treatments to "catch up."
4. Claim 3: Balancing the rate of convergence implies efficiency.
   This follows from the rate-optimal allocation discussed before.

Application: Agricultural extension service for farmers in Odisha, India

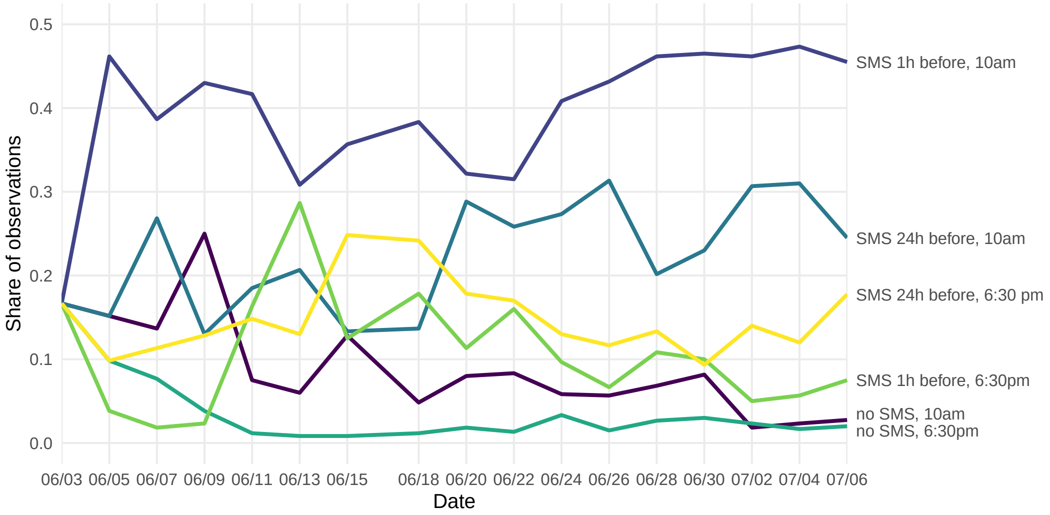
- Odisha (India), 2019.
  NGO Precision Agriculture for Development,
  and Government of Odisha.
- **Context**: Enrolling rice farmers into customized advice service by mobile phone.

  > *[...] to build, scale, and improve mobile phone-based agricultural extension with the goal of increasing productivity and income of 100 million smallholder farmers and their families around the world.*

- Sample: 10,000 calls,
  divided into waves of 600.
- 6 **treatments**:
  - The call is pre-announced via SMS 24h before, 1h before, or not at all.
  - For each of these, the call time is either 10am or 6:30pm.
- **Outcome**: Did the respondent answer the enrollment questions?

Odisha

Assignment shares over time



Outcomes and posterior parameters

Treatment		Outcomes			Posterior		
Call time	SMS alert	m_T^d	r_T^d	r_T^d/m_T^d	mean	SD	p_T^d
10am	-	903	145	0.161	0.161	0.012	0.009
10am	1h ahead	3931	757	0.193	0.193	0.006	0.754
10am	24h ahead	2234	400	0.179	0.179	0.008	0.073
6:30pm	-	366	53	0.145	0.147	0.018	0.011
6:30pm	1h ahead	1081	182	0.168	0.169	0.011	0.027
6:30 pm	24h ahead	1485	267	0.180	0.180	0.010	0.126

m_T^d: Number of observations, r_T^d: Number of successes, $p_T^d = P_T\left(d = \text{argmax}_{d'}\, \theta^{d'}\right)$.

Minimizing squared error

Maximizing in-sample outcomes

Optimizing policy choice: Average outcomes

Outlook

- Utilitarian welfare
- Combinatorial allocation
- Testing in a pandemic

Conclusion and summary

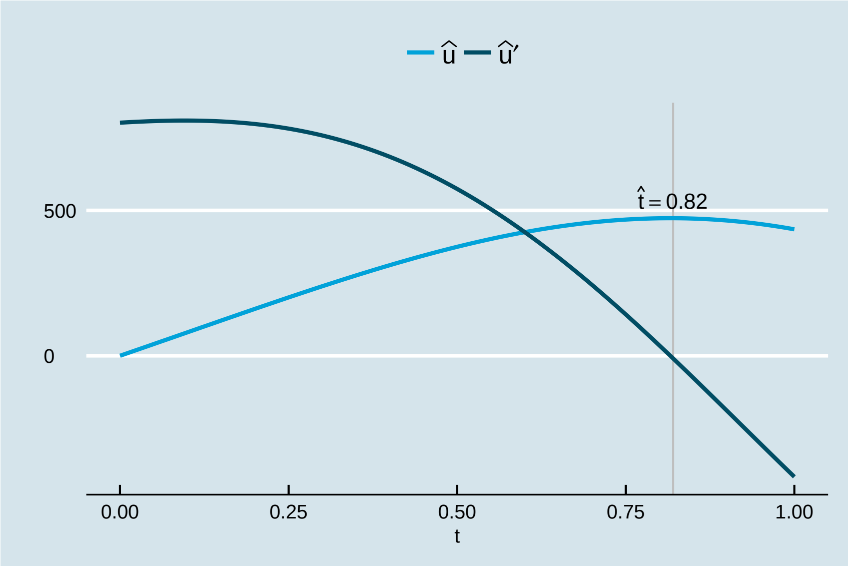
Maximizing utilitarian welfare

- For both in-sample regret and policy regret:
  Objectives are defined in terms of **observable outcomes**.
- Contrast this to welfare economics / optimal tax theory:
  Objectives are defined in terms of **revealed preference**.
- Quantification: **Equivalent variation**.
  What money transfer would make people indifferent to a given policy change?
- Operationalization through the **envelope theorem**:
  In assessing welfare effects, we can hold behavior constant.

Posterior expected social welfare (Kasy, 2019)

- Under standard assumptions of optimal taxation:
  Social welfare:
  $$u(t) = \lambda \int_0^t m(x)dx - t \cdot m(t),$$
  where λ is a welfare weight, $m(\cdot)$ is an average response, t is a tax rate.
- With experimental variation and Gaussian process prior:
  Posterior expected welfare:
  $$E[u(t)|data] = \boldsymbol{D}(t) \cdot \left[\boldsymbol{C} + \sigma^2 \boldsymbol{I}\right]^{-1} \cdot \boldsymbol{Y}.$$
- **Optimal tax rate**:
  $$\operatorname*{argmax}_t \; E[u(t)|data].$$

Example: RAND health insurance experiment, $\lambda = 1.5$



Experimental design problem

- Expected welfare after the experiment: $\max_t E[u(t)|data]$.
- Ex-ante expected welfare: $E[\max_t E[u(t)|\mathrm{data}]]$.
- Experimental design problem:

$$\underset{\mathrm{design}}{\operatorname{argmax}}\; E[\max_t E[u(t)|\mathrm{data}]].$$

  Maximize the expectation of a maximum of an expectation!
- If we allow for adaptivity:
  Additional layers of expectation and maximization for each wave.
  Numerically infeasible.

The knowledge gradient method

- Knowledge gradient method:
  An approximation successfully applied in the Bayesian optimization literature.
- Pretend that the experiment ends after the next wave. Solve

$$\underset{\text{assignment now}}{\operatorname{argmax}} \; E[\max_t E[u(t)|\text{data after this wave}]].$$

- This ignores the option-value of adapting in the future!
  But it provides an excellent approximation in practice.

Combinatorial allocation (Kasy and Teytelboym, 2020a)

Setup

- Select an **allocation** to maximize an objective, e.g.:
  - Allocate girls and boys across classrooms to max average test scores;
  - Allocate refugees across locations to max employment.
- Number of possible of allocations is potentially huge: exponential in number of possible matches and in batch size.
- Observe the outcome of each match (combinatorial semi-bandit).

Main result

- Prior-independent, finite-sample regret bound for Thompson algorithm that does *not* grow in batch size and grows only as $\sqrt{\# \text{ matches}}$.
- Thompson still achieves the efficient rate of convergence.

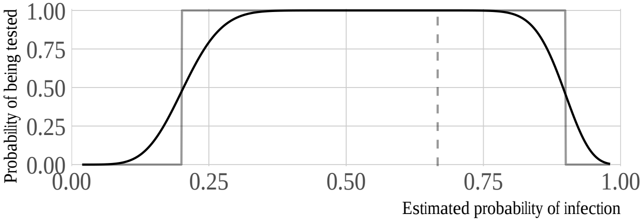
Testing in a pandemic (Kasy and Teytelboym, 2020b)

Setup

- Priority testing for symptomatic patients vs. random testing?
- How to optimally allocate costly disease-testing resources over time?
- Two costly errors if we do not test an individual:
  - False quarantine—opportunity costs of work and social life;
  - False release—costs of potentially spreading the disease further.

Thompson

- Initial exploration, eventually testing individuals
  with an **intermediate** likelihood of being infected.



Conclusion

- Any decision problem requires specification of an objective.
- The choice of objective matters for experimental design.
- Some possible choices:
  1. Squared error of effect estimates.
  2. In-sample regret.
  3. Policy-regret.
  4. Utilitarian welfare for policy choice.
- I discussed simple algorithms targeting each of these objectives.

Algorithms for these objectives

1. Expected **squared error**: Minimize

$$\mathrm{Var}(\beta|\boldsymbol{X}) - \overline{C}' \cdot (C + \sigma^2 I)^{-1} \cdot \overline{C}.$$

2. **In-sample regret** and squared error: γ-Thompson, with assignment probabilities

$$(1-\gamma) \cdot p_t^{dx} + \gamma/k, \qquad\qquad p_t^d = P_t\left(d = \underset{d'}{\operatorname{argmax}}\ \theta^{d'}\right).$$

3. **Policy regret**: Exploration sampling, with assignment probabilities

$$q_t^d = S_t \cdot p_t^d \cdot (1 - p_t^d), \qquad\qquad S_t = \frac{1}{\sum_d p_t^d \cdot (1 - p_t^d)}.$$

4. **Utilitarian welfare**: Knowledge gradient method for social welfare,

$$\underset{\text{assignment now}}{\operatorname{argmax}} \quad E[\max_t E[u(t)|\text{data after this wave}]].$$

Summary of theoretical findings

1. **Randomization is sub-optimal** in general decision problems:
   Randomization never decreases achievable Bayes / minimax risk,
   and is strictly sub-optimal if the optimal deterministic procedure is unique.
2. **Measure of balance (MSE)**:
   The expected MSE of an assignment is a measure of balance,
   and can be minimized for optimal assignments for estimation.
3. γ-**Thompson sampling (In-sample regret and MSE)**:
   In-sample regret is asymptotically proportional to γ.
   The variance of treatment effect estimates is decreasing in γ.
4. **Exploration sampling (Policy regret)**:
   The oracle optimal allocation equalizes power across suboptimal treatments.
   Exploration sampling achieves this in large samples,
   and is thus (constrained) rate-efficient.

Web apps implementing the proposed procedures

- Minimizing expected squared error:
  `https://maxkasy.github.io/home/treatmentassignment/`
- Maximizing in-sample outcomes:
  `https://maxkasy.github.io/home/hierarchicalthompson/`
- Informing policy choice:
  `https://maxkasy.shinyapps.io/exploration_sampling_dashboard/`

Thank you!