Adaptive maximization of social welfare in theory and practice

Maximilian Kasy

October 2022

Introduction

How should a policymaker act,

- who aims to maximize social welfare,

  Weighted sum of utility.

  $\Rightarrow$ Tradeoff redistribution vs. cost of behavioral responses.

- and needs to learn agent responses to policy choices?

  Adaptively updated policy choices.

  $\Rightarrow$ Tradeoff exploration vs. exploitation.

Introduction

How should a policymaker act,

- who aims to maximize social welfare,

  Weighted sum of utility.

  $\Rightarrow$ Tradeoff redistribution vs. cost of behavioral responses.

- and needs to learn agent responses to policy choices?

  Adaptively updated policy choices.

  $\Rightarrow$ Tradeoff exploration vs. exploitation.

Introduction

How should a policymaker act,

- who aims to maximize social welfare,

  Weighted sum of utility.

  $\Rightarrow$ Tradeoff redistribution vs. cost of behavioral responses.

- and needs to learn agent responses to policy choices?

  Adaptively updated policy choices.

  $\Rightarrow$ Tradeoff exploration vs. exploitation.

Taxes and bandits

- **Optimal tax theory**
  - Mirrlees (1971); Saez (2001); Chetty (2009)
- **Multi-armed bandits**
  - Bubeck and Cesa-Bianchi (2012); Lattimore and Szepesvári (2020)
- This talk: **Merging bandits and welfare economics**.
  - Unobserved welfare, as in optimal taxation.
  - Unknown response functions (treatment effects), as in multi-armed bandits.

Roadmap

- Part I:
  - With **Nicolò Cesa-Bianchi and Roberto Colomboni**.
  - A minimal model of adaptive welfare maximization.
  - Lower and upper bounds on adversarial regret.
  - Comparison to related learning problems.
- Part II:
  - With **Frederik Schwertner**.
  - Design of an adaptive basic income experiment in Germany.
  - Building on our ongoing conventional RCT.
  - Algorithm:

    Structural model of labor supply.

    $\Rightarrow$ MCMC sample from posterior for parameters, social welfare.

    $\Rightarrow$ Adaptive assignment shares to policies.

Review: Optimal taxation

- Social welfare = weighted sum of individual utilities.
- Welfare weights:

  Relative value of a marginal lump-sum \$ across individuals.

  $\approx$ Distributional preferences (rich vs. poor, healthy vs. sick,...)
- Envelope theorem:
  - Behavioral responses to marginal tax changes don't affect individual utilities.
  - They only impact public revenue (absent externalities).

  $\Rightarrow$ Impact on revenue is a sufficient statistic.
- Absent income effects:

  Consumer surplus

  $=$ Equivalent variation

  $=$ integrated response function.

Review: Adversarial bandits

- Canonical bandit problems:
  - Assign treatment sequentially.
  - Observe previous outcomes before the next assignment.
- Regret:

  How much worse is an algorithm

  than the best alternative in a given comparison set (e.g., fixed treatments).
- Two approaches for analyzing bandits:
  1. Stochastic: Potential outcomes are i.i.d. draws from some distribution.
  2. Adversarial: Potential outcomes are an arbitrary sequence.
- Adversarial regret guarantees:
  - Bound regret for arbitrary sequences.
  - We can do that because the stable comparison set substitutes for the stable data generating process.

Setup: Tax on a binary choice

Each time period $i = 1, 2, \dots, T$:

- Policymaker (algorithm):
  - Chooses tax rate $x_i \in [0, 1]$.
- Agent i:
  - Willingness to pay: $v_i \in [0, 1]$.
  - Response function: $G_i(x) = \mathbf{1}(x \leq v_i)$
  - Binary agent decision: $y_i = G_i(x_i)$.
- Observability:
  - After period i, we observe y_i.
  - We do *not* observe welfare $U_i(x_i)$.

Social welfare

Weighted sum of public revenue and private welfare:

$$U_i(x_i) = \underbrace{x_i \cdot \mathbf{1}(x_i \leq v_i)}_{\text{Public revenue}} \quad + \quad \lambda \cdot \underbrace{\max(v_i - x_i, 0)}_{\text{Private welfare}}.$$

We can rewrite private welfare as an integral (consumer surplus):

$$U_i(x) = \underbrace{x \cdot G_i(x)}_{\text{Public revenue}} \quad + \quad \lambda \cdot \underbrace{\int_x^1 G_i(x')dx'}_{\text{Private welfare}}.$$

Cumulative demand, welfare and regret

- Cumulative demand:

$$\mathbb{G}_T(x) = \sum_{i \leq T} G_i(x).$$

- Cumulative welfare for a constant policy x:

$$\mathbb{U}_T(x) = \sum_{i \leq T} \mathbb{U}_i(x) = x \cdot \mathbb{G}_T(x) + \lambda \int_x^1 \mathbb{G}_T(x')dx'.$$

- Cumulative welfare for the policies x_i actually chosen:

$$\mathbb{U}_T = \sum_{i \leq T} \mathbb{U}_i(x_i).$$

- Adversarial regret:

$$\mathcal{R}_T(\{v_i\}_{i=1}^T) = \sup_x E\left[\mathbb{U}_T(x) - \mathbb{U}_T \middle| \{v_i\}_{i=1}^T\right].$$

The structure of observability

Choice $\boldsymbol{x_i}$ reveals $\boldsymbol{G_i(x_i)}$. But

$$U_i(x) - U_i(x') = \left[x \cdot G_i(x) - x' \cdot G_i(x')\right] + \lambda \int_x^{x'} G_i(x'')dx''$$

depends on values of $G_i(x'')$ for $x'' \in [x, x']$!

Different from standard adaptive decision-making problems:

- Multi-armed bandits:
  Observe welfare for the choice made.

- Online learning:
  Observe welfare for all possible choices.

- Online convex optimization:
  Observe gradient of welfare for the choice made.

Lower bound on regret

Theorem

There exists a constant $C > 0$ such that,
for any algorithm for the choice of $x_1, x_2, \ldots$
and any time horizon $T \in \mathbb{N}$:

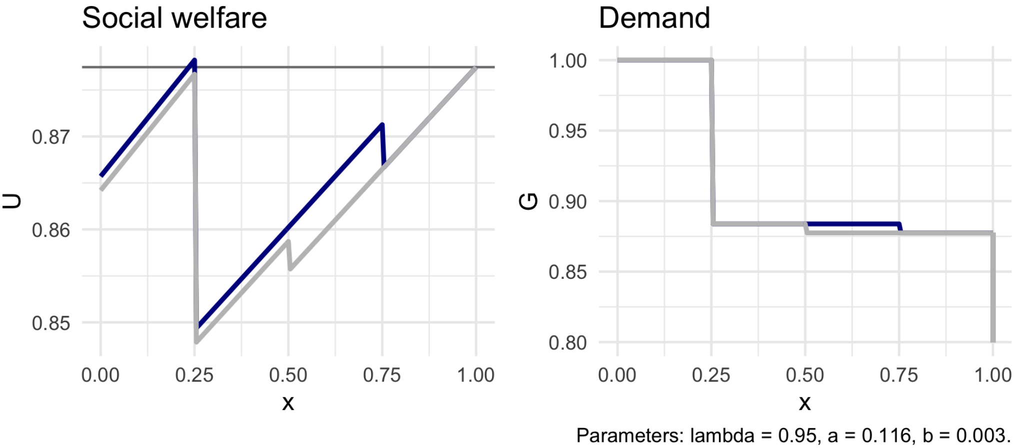
There exists a sequence $(v_1, \ldots, v_T)$ for which

$$\mathcal{R}_T(\{v_i\}_{i=1}^T) \geq C \cdot T^{2/3}.$$

Sketch of proof: Lower bound on regret

- Stochastic regret $\leq$ adversarial regret.
  (Since average $\leq$ maximum.)
- Construct a distribution for v with 4 points of support, e.g. $(\frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1)$.
- Choose the probability of each of these points such that
  1. The two middle points are far from optimal.
  2. Learning which of the two end points is optimal requires sampling from the middle.
     (Because of the integral term.)

Construction for the proof of the lower bound



Tempered Exp3 for social welfare

Require: Tuning parameters K, γ and η.

Set $\tilde{x}_k = (k-1)/K$, initialize $\hat{\mathbb{G}}_k = 0$ for $k = 1, \ldots, K+1$.

for individual $i = 1, 2, \ldots, T$ **do**

  for gridpoint $k = 1, 2, \ldots, K+1$ **do**

    Set

$$\widehat{\mathbb{U}}_{ik} = \tilde{x}_k \cdot \widehat{\mathbb{G}}_{ik} + \tfrac{\lambda}{K} \cdot \sum_{k'>k} \widehat{\mathbb{G}}_{ik'}, \quad p_{ik} = (1-\gamma) \cdot \frac{\exp(\eta \cdot \widehat{\mathbb{U}}_{ik})}{\sum_{k'} \exp(\eta \cdot \widehat{\mathbb{U}}_{ik'})} + \frac{\gamma}{K+1}.$$

  end for

  Choose k_i at random according to the probability distribution $(p_1, \ldots, p_{K+1})$.

  Set $x_i = \tilde{x}_{k_i}$, and query y_i accordingly.

  Update

$$\hat{\mathbb{G}}_{k_i} = \hat{\mathbb{G}}_{k_i} + \frac{y_i}{p_{ik_i}}.$$

end for

Upper bound on regret

Theorem

Consider the algorithm "Tempered Exp3 for social welfare."
There exists a constant C' and choices for K, γ, η such that,
for any sequence $(v_1, \ldots, v_T)$,

$$\mathcal{R}_T(\{v_i\}_{i=1}^T) \leq C' \cdot \log(T)^{1/3} \cdot T^{2/3}.$$

Note:

- Same rate as the lower bound, up to the logarithmic term.
- Upper bounds on adversarial regret
  are closely related to "Blackwell approachability."

Sketch of proof: upper bound on regret

- Discretize to balance the approximation error against the cost of having to learn $\mathbb{G}_i$ on more points.
- $\widehat{\mathbb{G}}$ is an unbiased estimator for cumulative demand $\mathbb{G}_i$.
  $\widehat{\mathbb{U}}$ is an unbiased estimator for cumulative discretized welfare.
- Consider $W_i = \sum_k \exp(\eta \cdot \widehat{\mathbb{U}}_{ik})$.
  - $E[\log W_T]$ is bounded below by η times optimal constant policy welfare.
  - $E\left[\log\left(\frac{W_i}{W_{i-1}}\right)\right]$ is bounded above by a combination of expected $\mathbb{U}_i$, and a term based on the second moment of $\widehat{\mathbb{U}}_i$.
- Bounding this second moment, and optimizing tuning parameters, yields the bound on adversarial regret.

Comparison to related learning problems

- **Monopoly pricing**:
  - Monopolist profits:

$$U_i^{MP}(x) = \underbrace{x \cdot G_i(x).}_{\text{Monopolist revenue}}$$

  - Easier – like a continuous multi-armed bandit.

- **Bilateral trade**:
  - Buyer plus seller welfare:

$$U_i^{BT}(x) = G_i^b(x) \cdot \underbrace{\int_0^x G_i^s(x')dx'}_{\text{Seller welfare}} + G_i^s(x) \cdot \underbrace{\int_x^1 G_i^b(x')dx'}_{\text{Buyer welfare}}.$$

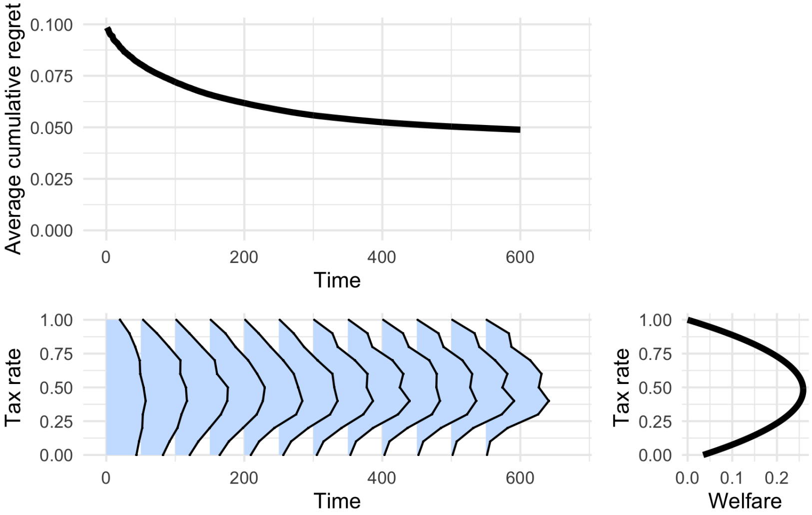
  - Harder – even gradients depend on global information.

Comparison of regret rates

Model	Policy space		Objective function	
	Discrete	Continuous	Pointwise	One-sided Lipschitz
Monopoly price setting	$T^{1/2}$	$T^{2/3}$	Yes	Yes
Optimal tax	$T^{2/3}$	$T^{2/3}$	No	Yes
Bilateral trade	$T^{2/3}$	T	No	No

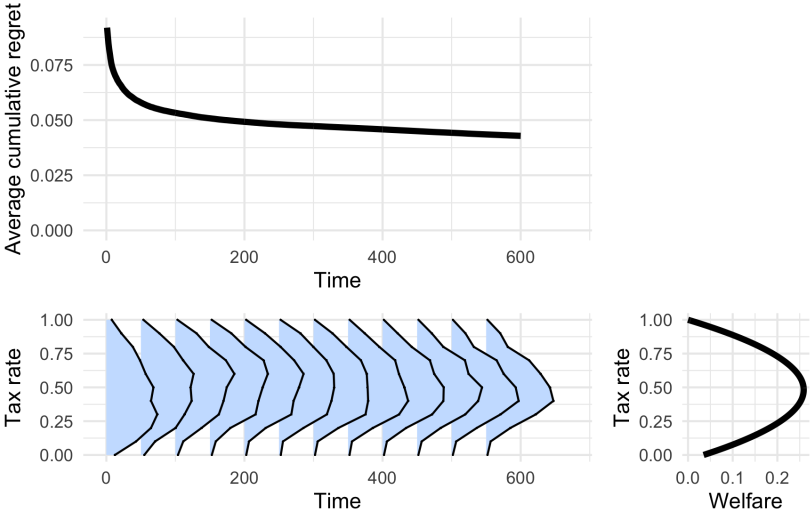
- Rates are up to logarithmic terms.
- They reflect:
  1. Information structures:
     Pointwise (like bandit) vs. global (require exploration away from optimum).
  2. Smoothness properties:
     One-sided Lipschitzness allows us to bound the discretization error.

Algorithm performance for $v \sim U[0, 1]$



1000 simulation repetitions. alpha = 1, beta = 1, K = 10, lambda = 0.7

Time-dependent tuning parameters

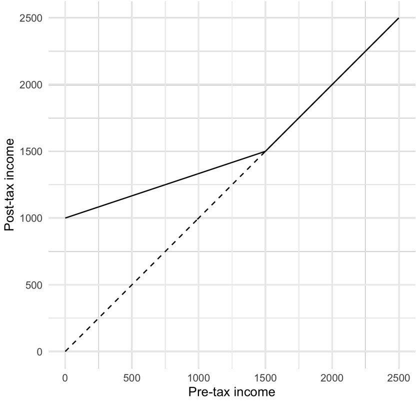


1000 simulation repetitions. alpha = 1, beta = 1, K = 10, lambda = 0.7

In the field: An adaptive basic income experiment in Germany

- Currently:
  - Classic RCT.
  - Evaluating a basic income (lump sum).
  - With the NGO “Mein Grundeinkommen” in Germany.
- In preparation:
  - Adaptive follow-up.
  - Negative income tax:
    Basic income y_0,
    taxed away until 0 transfer is reached.
  - Net-of-tax rate w.

Example: $(y_0, w) = (1000, 1/3)$



Policy grid

Basic income y_0, net-of-tax rate w

(0,0)	–	–	–
–	(500, 1/4)	(500, 1/2)	(500, 3/4)
–	(1000, 1/4)	(1000, 1/2)	(1000, 3/4)
–	(1500, 1/4)	(1500, 1/2)	(1500, 3/4)

- Every 6 months, a new cohort of participants will be enrolled.
- Participants receive basic income for 12 months.
- Fixed number of observations in the control group (0,0).
- Assignment shares across the 9 policy combinations are updated across waves.

Algorithm construction for the basic income experiment

1. Structural model of labor supply:
   - Extensive and intensive margins.
   - Non-convex budget sets.
   - Observations near kink $\Rightarrow$ optimization errors.
   - Observed and unobserved heterogeneity.
2. MCMC (Metropolis-Hastings):

   Sample from the posterior for structural parameters.

   $\Rightarrow$ Posterior distribution of social welfare for policy choices.

   $\Rightarrow$ Posterior probability that a policy is optimal.
3. Tempered Thompson sampling:
   - Like tempered Exp3.
   - But with "probability optimal" replacing the Exp3 term.

Structural model of labor supply

- Individual utility:

$$u_i(y) = \underbrace{y - T(y)}_{\text{Consumption}} - \underbrace{\frac{y}{\beta}[\log(y) - 1 - \alpha_i]}_{\text{Disutility of work}} - \underbrace{\left(\frac{\exp(\alpha_i)}{\beta} + \eta_i\right)}_{\text{Fixed cost of working}} \cdot \mathbf{1}(y > 0),$$

- where
  - $y \geq 0$ is reported earnings,
  - $T(y)$ is net taxes owed,
  - α_i shifts the intensive margin,
  - η_i shifts the extensive margin.

Labor supply and welfare for linear budget sets

- Linear tax schedule:
$$y - T(y) = y_0 + wy.$$
- FOC for labor supply, conditional on $y > 0$:
$$w = \frac{\log(y)}{\beta} - \frac{\alpha_i}{\beta}.$$
- Thus
$$y_i = \underbrace{\exp(\alpha_i + \beta w),}_{\text{Labor supply conditional on } y_i > 0}$$
$$u_i = y_0 + \underbrace{\exp(\alpha_i) \cdot \frac{\exp(\beta w) - 1}{\beta} - \eta_i}_{\text{Net utility of working.}}.$$
- If net utility of working < 0, then $y_i = 0$ and $u_i = y_0$.

Negative income tax

- Individual has a choice between 3 options:
  0. Not working: $y = 0$;
  1. Working under basic income y_0, plus tax with net-of tax rate w;
  2. Working under $y_0 = 0$ and $w = 1$.
- Utilities of these 3 options

$$
\begin{aligned}
u_i^0 &= y_0, \\
u_i^1 &= y_0 + \exp(\alpha_i) \cdot \frac{\exp(\beta w) - 1}{\beta} - \eta_i, \\
u_i^2 &= \exp(\alpha_i) \cdot \frac{\exp(\beta) - 1}{\beta} - \eta_i.
\end{aligned}
$$

Completing the model

- Problems with this model:
  1. No probability mass near kink, discontinuous distribution of $\boldsymbol{y_i}$.
  2. Discontinuous likelihood as function of β.

  $\Rightarrow$ Breaks maximum likelihood and MCMC.
- Solution: Optimization error.
  - When choosing which of the two schedules to optimize for, agents observe α_i with (small) error ϵ_i. Then they choose optimally.

  $\Rightarrow$ Smooth distribution, likelihood.
- Parametric specification: Covariates $\boldsymbol{x}$,

$$
\begin{aligned}
\alpha|x &\sim N(x \cdot \gamma_\alpha, \sigma^2), \\
\eta|\alpha, x &\sim N\left(-x \cdot \gamma_\eta/\tau, 1/\tau^2\right) \\
\epsilon|\eta, \alpha, x &\sim N(0, \rho^2).
\end{aligned}
$$

Markov Chain Monte Carlo sampling from the posterior

- Metropolis-Hastings:
  - Proposal $\tilde{\theta}_{t+1} \sim N(\hat{\theta}_t, \Omega)$.
  - Acceptance of proposal based on $U_t \sim U([0,1])$, posterior π,

$$\hat{\theta}_{t+1} = \begin{cases} \tilde{\theta}_{t+1} & U_t \leq \pi(\tilde{\theta}_{t+1}) \big/ \pi(\hat{\theta}_t), \\ \hat{\theta}_t & \text{else.} \end{cases}$$

  - π is the stationary distribution of this Markov chain.
- Convergence requires careful tuning:
  - Optimal proposal distribution for a normal posterior (Rosenthal, 2011):

$$\Omega = \frac{(2.38)^2}{d} \cdot \Sigma,$$

    where Σ is the posterior variance, $d = \dim(\theta)$.

$\Rightarrow$ We estimate Σ via the Hessian $-\nabla^2\pi$ at $\operatorname{argmax} \pi$ (maximum a posteriori).

Tempered Thompson sampling

- Thompson sampling:
  - Assign treatment arm $\boldsymbol{x}$ with probability $\boldsymbol{P_i(X_i = x)}$ equal to
  - the posterior probability that $\boldsymbol{x}$ is optimal,

$$P_i\left(x = \underset{x' \in \mathcal{X}}{\operatorname{argmax}}\ \boldsymbol{U}(x')\right).$$

  $\Rightarrow$ Optimal convergence rate of regret (Agrawal and Goyal, 2012) for canonical bandits.
  - But too little exploration for welfare maximization.

- Tempered Thompson sampling:

$$\boldsymbol{P_i(X_i = x)} = (1-\gamma) \cdot \boldsymbol{P_i}\left(x = \underset{x' \in \mathcal{X}}{\operatorname{argmax}}\ \boldsymbol{U}(x')\right) + \frac{\gamma}{|\mathcal{X}|}.$$

  - The posterior probability that $\boldsymbol{x}$ is optimal takes the place of the exponential weights in the Tempered Exp3 algorithm.

Conclusion

- A canonical economic problem:

  Choosing policies to maximize social welfare,

  while needing to learn behavioral responses.

- More difficult than canonical bandits, monopoly pricing:

  Learning the optimal policy

  requires exploration of sub-optimal policies.

- Broader agenda:

  1. Adapt tools from machine learning for the purpose of public good.
     (Vs. profit maximization – monopoly pricing, ad click maximization...)
  2. Unify insights from (welfare) economics and computer science.
  3. Span the range from theoretical performance guarantees
     to practical implementation.

Thank you!