Adaptive maximization of social welfare

Maximilian Kasy

June 2022

Introduction

How should a policymaker act,

- who aims to maximize social welfare,

  Weighted sum of utility.

  ⇒ Tradeoff redistribution vs. cost of behavioral responses.

- and needs to learn agent responses to policy choices?

  Adaptively updated policy choices.

  ⇒ 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 responses, as in multi-armed bandits.

Co-authors

- *Nicolò Cesa-Bianchi and Roberto Colomboni*,
  for the theory of adversarial and stochastic
  lower and upper bounds on regret.
- *Frederik Schwertner*,
  for implementation of an adaptive basic income experiment in Germany.

Setup

Lower and upper bounds on regret

Comparison to related learning problems

Simulations

In the field: An adaptive basic income experiment in Germany

Setup: Tax on a binary choice

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

- One agent with willingness to pay $v_i \in [0, 1]$.
- Choices:
  - Tax rate $x_i \in [0, 1]$.
  - Individual 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):

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

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 $\boldsymbol{G_i(x'')}$ for $\boldsymbol{x''} \in [x, x']$!

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.

Setup

Lower and upper bounds on regret

Comparison to related learning problems

Simulations

In the field: An adaptive basic income experiment in Germany

Lower bound on regret

Theorem

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

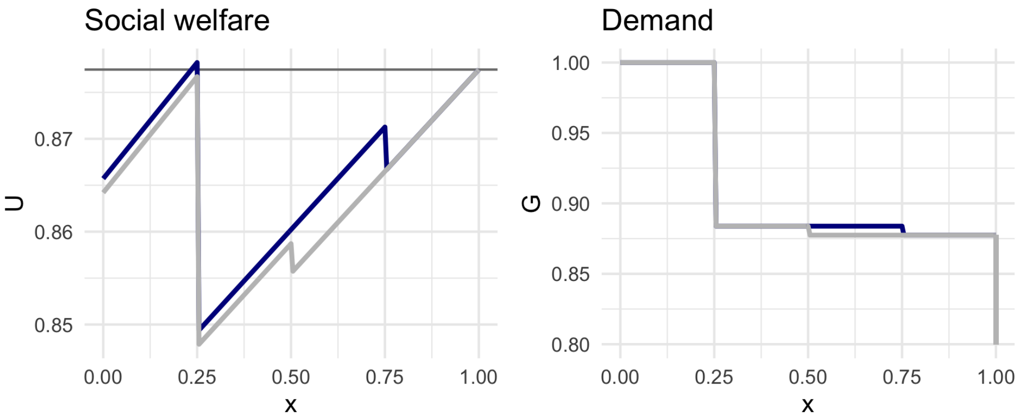
There exists a sequence $(v_1, \ldots, v_T)$ *for which*

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

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



Parameters: lambda = 0.95, a = 0.116, b = 0.003.

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 $(v_1, \ldots, v_T)$,

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

$\Rightarrow$ Same rate as the lower bound, up to the logarithmic term!

Sketch of proof

Setup

Lower and upper bounds on regret

Comparison to related learning problems

Simulations

In the field: An adaptive basic income experiment in Germany

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	Continuous	Discrete
Monopoly price setting	$T^{2/3}$	$T^{1/2}$
Optimal tax	$T^{2/3}$	$T^{2/3}$
Bilateral trade	T	$T^{2/3}$

- Rates are up to logarithmic terms.
- They reflect the different information structures in the three problems.

Setup

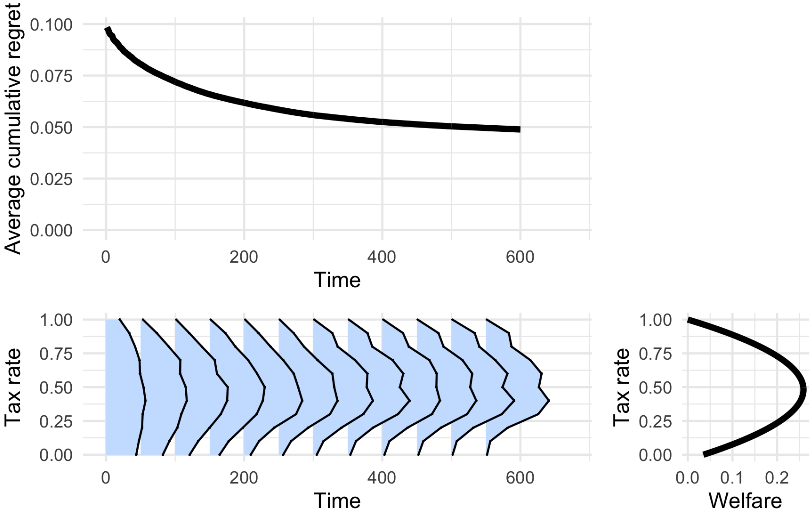
Lower and upper bounds on regret

Comparison to related learning problems

Simulations

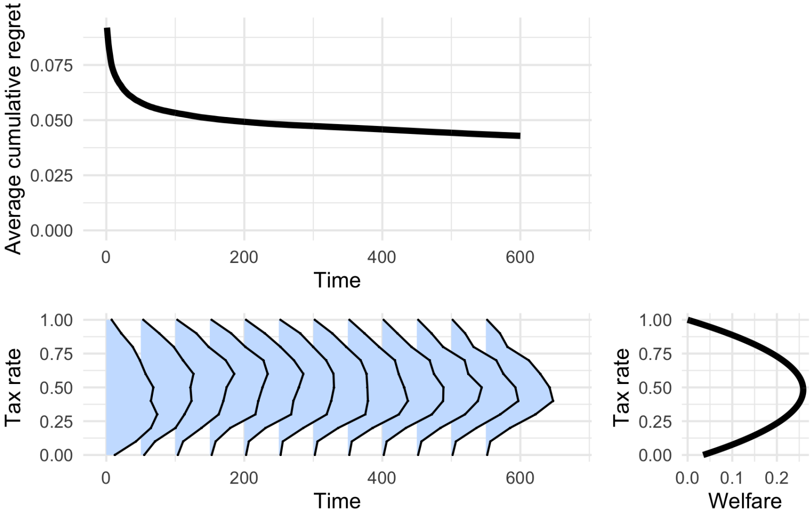
In the field: An adaptive basic income experiment in Germany

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

Setup

Lower and upper bounds on regret

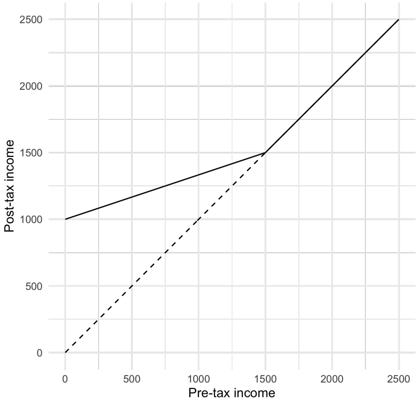
Comparison to related learning problems

Simulations

In the field: An adaptive basic income experiment in Germany

In the field: An adaptive basic income experiment in Germany

- Currently:
  Classic RCT evaluating a basic income, with the NGO "Mein Grundeinkommen" in Germany.
- In preparation: Adaptive follow-up.
  - Negative income tax: Basic income, taxed away until 0 transfer is reached.
  - Two policy parameters:
    Transfer size and tax rate.
    $\Rightarrow$ Grid of possible combinations.



Algorithm construction for the basic income experiment

- Structural model of labor supply:
  - Extensive and intensive margins.
  - Non-convex budget sets.
  - Measurement / optimization errors.
  - Observed and unobserved heterogeneity.
- Use MCMC (Metropolis-Hastings) to sample from the posterior for structural parameters.
- Map this into the posterior distribution of social welfare differences across policy choices.
- Assign policies using a version of tempered Thompson sampling.

Thank you!

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 an 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.