Adaptive maximization of social welfare

Maximilian Kasy

January 1, 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 responses, as in multi-armed bandits.

Co-authors

- *Dennis Hein*,
  for Thompson sampling with Gaussian process priors
  and random Fourier features.
- *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

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 $x_i \in [0, 1]$.
  - Binary agent decision $y_i = \mathbf{1}(x_i \leq v_i)$.
- Social welfare:
  - Public revenue $+ \lambda\cdot$ private welfare,

$$u_i(x_i) = x_i \cdot \mathbf{1}(x_i \leq v_i) + \lambda \cdot \max(v_i - x_i, 0).$$

- Observability:
  - After period i, we observe y_i.
  - We do *not* observe welfare $u_i(x_i)$.

Consumer surplus and cumulative social welfare

- Individual demand function: $G_i(x) = \mathbf{1}(x \leq v_i)$.
  Cumulative demand: $\bar{G}_T(x) = \sum_{i \leq T} G_i(x)$.
- We can rewrite private welfare as an integral (consumer surplus):

$$u_i(x) = x \cdot G_i(x) + \lambda \int_x^1 G_i(x')dx'.$$

- Cumulative welfare for a constant policy x:

$$U_T(x) = \sum_{i \leq T} u_i(x) = x \cdot \bar{G}_T(x) + \lambda \int_x^1 \bar{G}_T(x')dx'.$$

- Cumulative welfare for the policies x_i actually chosen:

$$U_T = \sum_{i \leq T} u_i(x_i).$$

The structure of observability

Recall: $u_i(x) = x \cdot G_i(x) + \lambda \int_x^1 G_i(x')dx'$.

- Choice x_i reveals $G_i(x_i)$.
- But $u_i(x)$ depends on values of $G_i(x')$ for $x' \in [x, 1]$!

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

In the field: An adaptive basic income experiment in Germany

Lower bound on stochastic and adversarial regret

Theorem

There exists a constant $C > 0$ *such that for any algorithm for the choice of* x_i*:*

1. *There exists a distribution of* v_i *such that* $\sup_x E\left[U_T(x) - U_T\right]$ *equals at least* $C \cdot T^{2/3}$.
2. *There exists a sequence* $(v_1, \ldots, v_T)$ *such that* $\sup_x E\left[U_T(x) - U_T \middle| \{v_i\}_{i=1}^T\right]$ *equals at least* $C \cdot T^{2/3}$.

Compare to the lower bound for stochastic / adversarial bandits: $C \cdot T^{1/2}$.

Sketch of proof

Tempered Exp3 for social welfare

Require: Tuning parameters K, γ and η.

Set $\tilde{x}_k = (k-1)/K$, initialize $\hat{G}_k = 0$ for $k = 1, \ldots, K+1$.
for individual $i = 1, 2, \ldots, T$ **do**
  for gridpoint $k = 1, 2, \ldots, K+1$ **do**
    Set

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

  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

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

end for

Adversarial upper bound

Conjecture

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

$$\sup_x E\left[U_T(x) - U_T \middle| \{v_i\}_{i=1}^T\right]$$

equals at most $C' \cdot T^{2/3} \cdot \log(T)$.

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

Sketch of proof

Setup

Lower and upper bounds on regret

In the field: An adaptive basic income experiment in Germany

In the field: An adaptive basic income experiment in Germany

- We are currently running a classic RCT evaluating a basic income with the NGO “Mein Grundeinkommen” in Germany.
- An adaptive follow-up is in preparation:
  - Negative income tax – basic income, taxed away until 0 transfer is reached.

  $\Rightarrow$ Two policy parameters: Transfer size and tax rate.
  We will focus on a small grid of possible combinations.
- Theoretical challenges:
  1. Multi-dimensional policies.
  2. Preferences with income effects.
  3. Avoiding tuning parameters.
  4. Exploiting smoothness, convexity.
- Practical challenge:

  This will be expensive...

Thank you!

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.)

Sketch of proof: upper bound on regret

- Discretize to balance the approximation error
  against the cost of having to learn $\bar{G}_i$ on more points.

- $\hat{G}$ is an unbiased estimator for cumulative demand $\bar{G}_i$.
  $\hat{U}$ is an unbiased estimator for cumulative discretized welfare.

- Consider $W_i = \sum_k \exp(\eta \cdot \hat{U}_{ik})$.
  - $E[\log W_T]$ 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 u_i,
    and a term based on the second moment of $\hat{u}_i$.

- Bounding this second moment, and optimizing tuning parameters,
  yields the bound on adversarial regret.