Inferring players' values from their actions

Xiang Yan

Harvard University

Nov 21, 2019

Content

- Example of inferring values in a 2x2 game
- Equilibrium-based players assumptions
- No-regret players assumption
- Quantal regret players assumption
- Application to ad auctions

Based on:

Stationary Concepts for Experimental 2x2-Games by Selten and Chmura

*Econometrics for Learning Agent*s by Nekipelov et, al.

A "quantal regret" method for structural econometrics in repeated games by Nisan and Noti

Estimation Task in a 2x2 Game

- How to infer x and y if we know how the game is played?
  - No dominant strategy

- e.g. $p_U = 0.11, q_L = 0.68$

- Assume Nash Equilibrium:
  - $q_L = 10/(x+1)$
  - $x \approx 13.7$

	Left	Right
Up	x, y	0, 18
Down	9, 9	10, 8

Equilibrium-based Model

- NE:

$$p_U = \frac{\mathrm{d_D}}{d_U + d_D},$$

$$q_L = \frac{c_R}{c_L + c_R}$$

	L	R
U	$a_L + c_L$ / b_U	a_R / $b_U + d_U$
D	a_L / $b_D + d_D$	$a_R + c_R$ / b_D

U: up D: down
L: left R: right

Player 1's payoff in the upper-left corner
Player 2's payoff in the lower-right corner

$a_L, a_R, b_U, b_D \geqslant 0$
$c_L, c_R, d_U, d_D > 0$

- In practice, NE is hard for players to compute
  - In the experiments, players are informed with their payoff after each play and their opponents' actions and total payoffs after 200 repeats.

Equilibrium-based Model

- Quantal Response Equilibrium
- For player 1 (row)
  - $E_U(q_L), E_D(q_L)$: expected payoff against q_L
  - $p_U = \frac{e^{\lambda E_U(q_L)}}{e^{\lambda E_U(q_L)} + e^{\lambda E_D(q_L)}}$
- Similarly
  - $q_L = \frac{e^{\lambda E_L(p_U)}}{e^{\lambda E_L(p_U)} + e^{\lambda E_R(p_U)}}$

- Solve the equation and fit λ through data

	L	**R**
U	$a_L + c_L$, b_U	a_R, $b_U + d_U$
D	a_L, $b_D + d_D$	$a_R + c_R$, b_D

Equilibrium-based Model

- Action-sampling Equilibrium
- Player 1: n samples where L played k times
  - Choose his best response according to samples
  - $\alpha_U(k) = 1 \text{ if } \frac{k}{n} > \frac{c_R}{c_L + c_R}; \frac{1}{2} \text{ if ” = ”; 0 else}$
  - $p_U = \sum_{k=0}^{n} \binom{n}{k} q_L^k (1 - q_L)^{n-k} \alpha_U(k)$
- Player 2: n samples where U played m times
  - $\alpha_L(m) = 1 \text{ if } \frac{m}{n} > \frac{d_U}{d_U + d_D}; \frac{1}{2} \text{ if “ = ”; 0 else}$
  - $q_L = \sum_{m=0}^{n} \binom{n}{m} p_U^{n-m} (1 - p_U)^m \alpha_L(m)$

	L	***R***
U	$a_L + c_L$, b_U	a_R, $b_U + d_U$
D	a_L, $b_D + d_D$	$a_R + c_R$, b_D

Equilibrium-based Model

- Payoff-sampling Equilibrium
- Player 1: n samples in total, L played k_U (k_D) times when he plays U (D)
  - Payoff in samples:
    - $\mathrm{H}_U = k_U(a_L + c_L) + (n - k_U)a_R$
    - $\mathrm{H}_D = k_D a_L + (n - k_D)(a_R + c_R)$
  - Play the one with higher payoff
    - $\beta(k_U, k_D) = 1$ if $H_U > H_D$; $\frac{1}{2}$ if " = "; 0 else
  - $p_U = \sum_{k_U=0}^{n} \sum_{k_D=0}^{n} \binom{n}{k_U} \binom{n}{k_D} q_L^{k_U+k_D} (1 - q_L)^{2n-k_U-k_D} \beta\,(k_U, k_D)$
- Similar for player 2

	L	***R***
U	$a_L + c_L$; b_U	a_R ; $b_U + d_U$
D	a_L ; $b_D + d_D$	$a_R + c_R$; b_D

Equilibrium-based Model

- Impulse Balance Equilibrium
- Security level: the maxmin payoff
  - $s_1 = \max[\min(a_L + c_L, a_R), \min(a_L, a_R + c_R)$
  - $s_2 = \max[\min(b_U, b_D + d_D), \min(b_U + d_U, b_D)]$
  - Second lowest payoff
- Transform the payoff matrix
  - Idea: a loss is twice weighted than a gain
  - Reduce the surplus payoff over s_i by a factor $\frac{1}{2}$

	L	***R***
U	$a_L + c_L$ b_U	a_R $b_U + d_U$
D	a_L $b_D + d_D$	$a_R + c_R$ b_D

8 ; 6^*	0 ; 14
7^* ; 7	10 ; 4

$\longrightarrow$

$7{,}5$; 6	0 ; 10
7 ; $6{,}5$	$8{,}5$; 4

	L	**R**
U	0 ; d_U^*	c_R^* ; 0
D	c_L^* ; 0	0 ; d_D^*

Equilibrium-based Model

- Impulse Balance Equilibrium
- Security level: the maxmin payoff
- Transform the payoff matrix
- Impulse balance equations:
  - $p_U q_R c_R^* = p_D q_L c_L^*$
  - $p_U q_L d_U^* = p_D q_R d_D^*$
- Then
  - $p_U = \frac{q_L c_L^*}{q_L c_L^* + (1-q_L) c_R^*}$, $q_L = \frac{(1-p_U) d_D^*}{p_U d_U^* + (1-p_U) d_D^*}$
  - $p_U = \frac{\sqrt{c}}{\sqrt{c}+\sqrt{d}}$, $q_L = \frac{1}{1+\sqrt{cd}}$, $c = \frac{c_L^*}{c_R^*}$, $d = \frac{d_U^*}{d_D^*}$

	L	***R***
U	*0* d_U^*	c_R^* *0*
D	c_L^* *0*	*0* d_D^*

Estimation Task in a 2x2 Game

- Infer one element of the payoff matrix through repeatedly playing samples

- Performance:
  1. Impulse balance equilibrium
  2. Payoff-sampling equilibrium
  3. Action-sampling equilibrium
  4. Quantal response equilibrium
  5. Nash equilibrium

	L	R
U	$a_L + c_L$; b_U	a_R ; $b_U + d_U$
D	a_L ; $b_D + d_D$	$a_R + c_R$; b_D

Estimation Task in a 2x2 Game

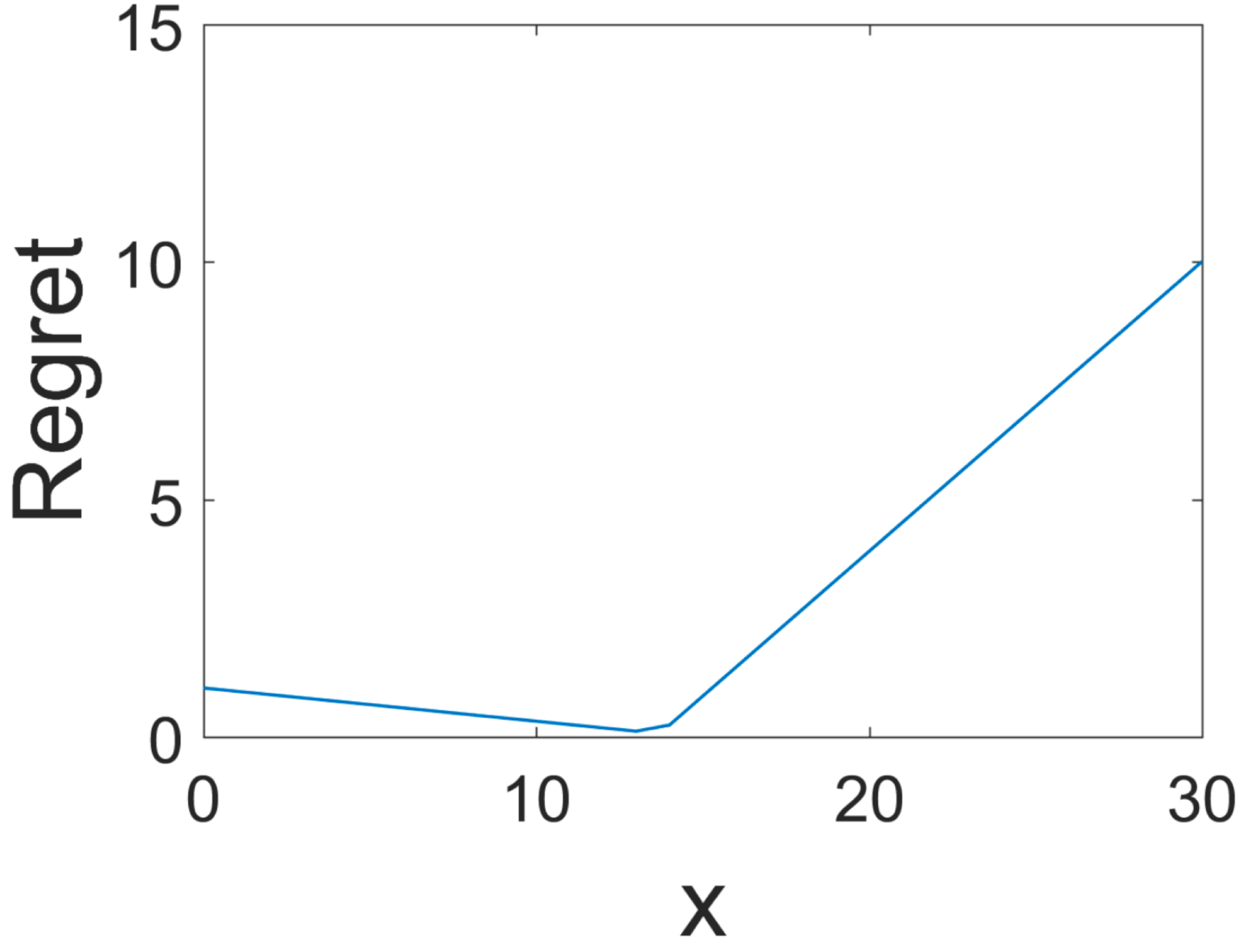
- Other equilibrium is still hard to compute.
- What if players are using no-regret learning?

$$regret_i(\theta, \vec{a}) = \frac{1}{T}(max_{a'_i \in A_i} \sum_{t=1}^{T} u_i(a'_i, a^t_{-i}, \theta) - \sum_{t=1}^{T} u_i(a^t_i, a^t_{-i}, \theta)).$$

  - An algorithm is a no-regret learning if the regret goes to 0 as T increase to infinity.

Estimation Task in a 2x2 Game

	Left	Right
Up	x, y	0, 18
Down	9, 9	10, 8



- $p_U = 0.11, q_L = 0.68$
- $x \approx 13$ for the lowest regret
- Recall NE leads to 13.7

Quantal Regret Model

- The prior of θ (i.e. x) is uniform distribution $p(\theta)$
- It is updated according to the regret:

$$\hat{\theta} = Z^{-1} \cdot \sum_{\theta} p(\theta) \cdot e^{-\lambda \cdot \sum_i regret_i(\theta, \vec{a})} \cdot \theta,$$

where

$$Z = \sum_{\theta} p(\theta) \cdot e^{-\lambda \cdot \sum_i regret_i(\theta, \vec{a})}$$

- λ : parameter reflecting how players relying on no-regret learning
  - If $\lambda = \infty$, it is equivalent to min-regret
- Performs better than equilibrium-based model and min-regret

Application to Auctions

- For given bidding $\boldsymbol{b}^t = (b_i^t, \boldsymbol{b}_{-i}^t), t = 1, \dots, T, (v, \epsilon)$ is no-regret pair if

$$\forall b' \in \mathbb{R}_+ : v \cdot \frac{1}{T} \sum_{t=1}^{T} \left(P_i^t(b', \mathbf{b}_{-i}^t) - P_i^t(\mathbf{b}^t) \right) \leq \frac{1}{T} \sum_{t=1}^{T} \left(C_i^t(b', \mathbf{b}_{-i}^t) - C_i^t(\mathbf{b}^t) \right) + \epsilon$$

where $P(\cdot)$ is winning probability, $C(\cdot)$ is the payment function

- Define deviation function:

$$\Delta P(b') = \frac{1}{T} \sum_{t=1}^{T} \left(P_i^t(b', \mathbf{b}_{-i}^t) - P_i^t(\mathbf{b}^t) \right) \qquad \Delta C(b') = \frac{1}{T} \sum_{t=1}^{T} \left(C_i^t(b', \mathbf{b}_{-i}^t) - C_i^t(\mathbf{b}^t) \right)$$

- Rationalizable Set $\mathcal{NR} = \{(v, \epsilon) \mid \forall b' \in \mathbb{R}_+ : v \cdot \Delta P(b') \leq \Delta C(b') + \epsilon \}$

Application to Auctions

- With some assumptions

ASSUMPTION 1. *The support of bids is a compact set* $B = [0, \bar{b}]$. *For each bidvector* $\mathbf{b^t_{-i}}$ *the functions* $P_i^t(\cdot, \mathbf{b}_{-i}^t)$ *and* $C_i^t(\cdot, \mathbf{b}_{-i}^t)$ *are continous, monotone increasing and bounded on* B.

ASSUMPTION 2. *For each* b_1 *and* $b_2 > 0$ *the incremental cost per click function*

$$ICC(b_2, b_1) = \frac{\Delta C(b_2) - \Delta C(b_1)}{\Delta P(b_2) - \Delta P(b_1)}$$

is continuous in b_1 *for each* $b_2 \neq b_1$ *and it is continuous in* b_2 *for each* $b_1 \neq b_2$. *Moreover for any* $b_4 > b_3 > b_2 > b_1$ *on* B: $ICC(b_4, b_3) > ICC(b_2, b_1)$.

ASSUMPTION 3.

(i) The rationalizable values per click are non-negative: $v \geq 0$.
(ii) There exists a global upper bound $\bar{\epsilon}$ *for the error of all players.*

Application to Auctions

- With some assumptions, the Hausdorf distance between the estimated $\widehat{\mathcal{NR}_B}$ and the real $\mathcal{NR}_B$ satisfies

$$d_H(\widehat{\mathcal{N}R}_B, \mathcal{N}R_B) \leq O((N^{-1}\log N)^{\gamma/(2\gamma+1)}), \quad \gamma = k + \alpha.$$

  where N is the number of samples. (B stands for bounded)
- It requires an estimation for $\Delta C(\Delta P^{-1}(\cdot))$ from data, which is easy.
- In stead of estimating v through $(v, \epsilon) \in \mathcal{NR}$ with smallest ϵ, do

$$\hat{v}_i = (\sum_j v_i^j \cdot e^{-\lambda\epsilon_i^j})/(\sum_j e^{-\lambda\epsilon_i^j})$$

Thank You